IOE, Faculty of Education and Society

UK Early Life Cohort Feasibility Study: Age 9-10 months sweep

User Guide (Version 1)

September 2025

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

Contact

Data queries: help@ukdataservice.ac.uk

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

Authors

Vilma Agalioti-Sgompou, Lisa Calderwood, Liam Curran, Pasco Fearon, Alissa Goodman, Nureen Mohamad Zaki, Alyce Raybould, Richard Silverwood, Konstantinos Tsigaridis, Erica Wong.

How to cite this guide

Agalioti-Sgompou V., Calderwood L., Curran, L., Fearon, P., Goodman, A., Mohamad Zaki, N., Raybould A., Silverwood R., Tsigaridis K., Wong E. (2025) *UK Early Life Cohort Feasibility Study: Age 9-10 months sweep 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 cls.ucl.ac.uk/data-access-training/citing-our-data/

Centre for Longitudinal Studies

Centre for Longitudinal Studies (CLS)

UCL Social Research Institute

University College London

20 Bedford Way, London WC1H 0AL

cls.ucl.ac.uk

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

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

Contents

About the Early Life Cohort Feasibility Study

The Early Life Cohort Feasibility Study (ELC-FS) collected data from parents on a new generation of UK-born babies in their first year of life. It captured information about their economic and social environments, and their health, wellbeing and development.

The main aim of the project was to test the feasibility of sampling and recruitment for an innovative new UK-wide birth cohort study.

The Early Life Cohort Feasibility Study drew a nationally representative sample of babies born in a two-month period from all four UK nations. In England, Wales and Scotland the babies were born in November and December 2022, and in Northern Ireland the babies were born in June and July 2023. Interviews took place when the babies were around 9-13 months old. The study was known to participants as ‘Generation New Era’.

1. Introduction

Data collection for the Early Life Cohort Feasibility Study (ELC-FS) took place between September 2023 and September 2024. The study was designed and managed by the UCL Centre for Longitudinal Studies (CLS) with support from a consortium of partners. Fieldwork was carried out by Ipsos. It was funded by the Economic and Social Research Council.

Ethical approval was provided the London-Central Research Ethics Committee (REC reference: 22-LO-0066).

The overarching aim of the Feasibility Study was to explore the feasibility of a new birth cohort study in the UK that would paint a nationally representative picture of the circumstances and lives of a new cohort of babies born at a critical time in the UK's history, and to understand how inequalities in early child development are changing over time.

The primary objective of the study was therefore to provide evidence on the potential for successful recruitment into a new national birth cohort study, and to inform on the best approach to design and measurement.

In order to accomplish this, the core aims were to:

- To draw a sample and recruit into a UK-representative study of babies in first year of life (around 9-10 months).
- To create an inclusive cohort, including families that are typically under-represented.
- To test feasibility of innovative measures, and linkages.
- To engage extensively to ensure scientific and policy value, public acceptability, and participant co-production.

The primary scientific aim of the study is to understand how inequalities in early child development are changing over time, and to learn whether the social and biological factors driving these trajectories are evolving. The data are of substantive value for identifying potential foci for early intervention and support.

ELC-FS was based on a sample of babies born in England, Wales and Scotland in November-December 2022 and in Northern Ireland in June-July 2023. In England, Wales and Scotland, samples were drawn from birth registration data that was matched with maternity records; samples were provided by NHS England, National Records Scotland and Public Health Scotland. In Northern Ireland, samples were drawn from maternity records only, provided by the Business Services Organisation on behalf of Health Trusts. Parents in each country had the option to opt-out of the study prior to a visit by an interviewer at their home.

The study aimed to recruit families when their babies were aged 8-10 months, with the target age of recruitment at 9 months; however, fieldwork sometimes occurred later than planned due to delays meaning the babies' ages ranged from 8-19 months, with most being between 9-13 months (70.7%). More information on the babies' ages and how this varies in different countries due to timing of fieldwork can be found in section 2.3 'Fieldwork dates'.

The main survey component involved interviews with parents (and the co-residential partners of parents, where applicable), including parents who live apart from their babies. There were three types of interview for each baby which were allocated following a doorstep screening exercise with an interviewer:

- The Primary Informant (PI) interview, for a parent living with their baby and who spends the most time caring for their baby (60-minute interview).

- The Own Household Parent (OHP) interview, for any parents living apart from their baby all or most of the time (40-minute interview). Details for these parents were in some cases available on the sample, and in some cases collected from Primary Informants during their interview.

- The Additional Informant (AI) interview, usually for a parent living with a Primary Informant, but also including the co-residential partners of Primary Informants or Own Household Parents who we refer to as AIOHPs (30-minute interview).

Interviews were usually carried out face-to-face, but telephone and Teams video-call interviews were also an option. The Additional Informant and Own Household

Parents interviews also had a web completion option. In some rare instances the Primary Informant also completed by web, but this was not explicitly offered. When the main fieldwork period was finished, non-responding households were sent a postal invitation to complete a 30-40 minute web survey (the online follow-up survey). **Mode of interview can be found in variable 'int_mode'**.

Interviews were completed with 3126 parents (1910 PIs, 1156 AIs (or which 4 were AIOHPs), 60 OHPs). One family with two parent interviews had triplets and are not included in the Safeguarded (End User License, EUL) dataset.

Overall, an interview being achieved in the child's main household (PI or AI interview) in 1918 families, with 1960 children. There were a further 15 families, with 16 children, where only interviews in the child's second household were achieved (i.e. only OHP or AIOHP interviews were achieved, and no PI or AI interviews in the child's main household). 18 respondents were interviewed among these 15 families. The total numbers of interviews in the child's main household included in the Safeguarded dataset (i.e. without the triplet family) are therefore 3106 parents, 1917 families and 1957 children.

The study response rate was 49% (the percent of families with at least one interview in the child's main household out of eligible study families issued to opt-out stage) and the survey response rate was 51% (the percent of families with at least one interview in the child's main household out of those who did not opt-out i.e. the families issued to field).

In total, 1933 families with 1976 children took part (1932 families with 1973 children in EUL):

- 1015 families in England (1014 in EUL)
- 279 in Wales
- 319 in Scotland
- and 320 in Northern Ireland.

There were 1891 families with one baby, 41 families with two babies (twins), and 1 family with three babies (triplets).

There were 741 families where only 1 parent took part, 1190 where 2 parents took part, and one where three parents took part.

1,853 mums and 1,272 dads took part (and one parent who responded ‘don’t know’ for the question about their sex at birth). This is 1852 mums and 1271 dads in the EUL version.

Study participants were asked to provide consent for completing a survey interview. In addition, a subsample of participants was asked to give consent for providing their own and their baby’s saliva sample so that DNA could be extracted for genetic research. **Participants who were asked for consent to provide saliva samples can be identified through the variable ‘sal_substudy’.** Two consent approaches to allow the ELC-FS study team to access information held in various administrative records were also tested. The saliva consent and data linkage consent data are not included as part of this data deposit.

This user guide provides information about the data arising from the Early Life Cohort Feasibility Study and accompanies the data shared via the UK Data Service. A full account of the study development and fieldwork procedures can be found in the Generation New Era Technical Report produced by Ipsos (publication forthcoming on the CLS website). The numbers of interviews, families and babies in the datasets and this user guide differ slightly from those included in the Technical Report. This is due to the data and user guide using the fieldwork data post validation by CLS, whereas the Technical Report uses Ipsos’s fieldwork data prior to validation by CLS.

In addition to this user guide, the ELC-FS documentation accompanying the data deposit includes the ELC-FS Questionnaire.

2. Fieldwork

2.1 Sample

2.1.1 Sample frame

The sampling frame was based on birth registrations linked to NHS maternity records in England, Wales and Scotland. Birth registrations provide universal coverage of the population of babies and contain key characteristics of the infant, mother and father, including fathers living apart from the baby (where jointly registered). In Northern Ireland, the sampling frame was based on maternity records; therefore, the sample included only the mother's contact details, did not list names or addresses for fathers and had much less information about the baby and mother.

Further detail on the sampling frame and relevant sample exclusions can be found in the Generation New Era Technical Report.

2.1.2 Sample design

Scotland, Wales and Northern Ireland were oversampled relative to England. The sample design for the feasibility study also included two additional boosts in England only: an ethnic minority boost of Black African and Black Caribbean babies and Pakistani and Bangladeshi babies, and an area-based low-income boost.

The samples were selected in two stages. At the first stage, a random sample of areas using Census geographies was selected for each country with probability proportionate to the number of births. At the second stage, a stratified random sample of children was sampled within each selected area. The approach to sampling at this second stage varied by country, as described below. Because of differences in the Census geographies across the four countries, the definitions of the areas varied, comprising merged lower super output areas (LSOAs) for England and Wales, data zones (DZs) for Scotland, and super output areas (SOAs) for Northern Ireland.

The approach to sampling also varied. For England and Wales, both stages of the sampling were carried out by Ipsos. For Scotland and Northern Ireland however Ipsos carried out the first stage of sampling and then sent the data holders counts of the number of children to sample in each area so that they could carry out the second stage of sampling.

The 'returned sample' comprises the named sample after both stages of sampling, with any data holder exclusions applied, including the National Data Opt-Out in England.

2.2 Two-stage recruitment design and issued sample

The study used a two-stage recruitment approach. After removing some cases from the returned sample due to address anomalies, those remaining were issued to the initial opt-out stage. In the initial 'opt-out' stage, parents first received a notification mailing informing them they had been selected to be part of the study, and to contact the study if they did not want to take part or receive further information. Details on the administration of this recruitment approach can be found in the Generation New Era Technical Report. In the second stage, those parents who had not opted out were issued to field (the 'issued sample') and sent an invitation mailing with more information about the study and informing them an interviewer would be visiting their address to find out if they'd like to take part, unless they contacted the study.

A total of 3,633 eligible families (or birth events, meaning a selected child or children if a multiple birth) were issued to interviewers for the main fieldwork period, following the opt-out stage.

This issued sample total excludes a) ineligible families removed from the sample, e.g., because the family had moved overseas, b) families removed due to sensitive circumstances, and c) cases affected by an interviewer who was found to have been falsifying parts of interviews. All interviews achieved by this interviewer were deleted, though some cases were reissued later to achieve a legitimate interview, and these interviews appear in the shared data. They are not, however, counted as part of the total number of eligible families issued to field as presented in the Technical Report

and for the purposes of calculating response rates. See chapter 3 below, for further information on this.

After the main fieldwork period, the households of eligible families who did not respond were invited to take part in a shorter online follow-up survey.

2.3 Fieldwork dates

The fieldwork took place at different times across the four countries, in part because of when the sample could be drawn in those countries, and in part because of delays to fieldwork (see figure 1).

In England, Wales, and Scotland, the sample was drawn from births in November-December 2022. The opt-out 'notification' mailings were issued in August 2023. Main fieldwork in England and Wales then commenced in September 2023 and continued through February 2024. The online follow-up survey was then issued in March 2024, running until end of April 2024. In Scotland, the main fieldwork period commenced in September 2023 but then had to be stopped, recommencing in late November 2023. Fieldwork in Scotland finished in April 2024, and the online follow-up survey ran between May and June 2024.

In Northern Ireland, due to delays obtaining sample frame approvals, the sample was drawn from a two-month birth window in June-July 2023. The opt-out mailing was sent by the data holder in March 2024, the main fieldwork ran from April-July 2024, and the online follow-up survey ran between August-September 2024 which marked the final end of all fieldwork in all countries.

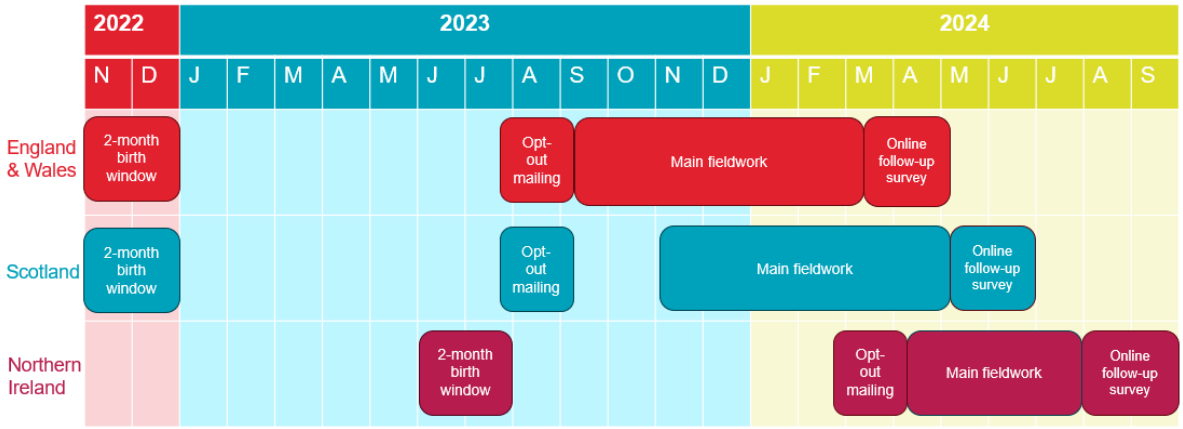


Figure 1: Study fieldwork dates across England & Wales, Scotland and Northern Ireland

This staggered fieldwork approach meant that babies had different average ages between the countries. In England and Wales, where fieldwork began when the oldest babies had just turned 10 months, babies were 8-19 months (mean age 11.6 months). In Scotland, fieldwork began when the oldest babies were 11 months, and babies were 11-20 months (mean age 13.9 months).

In Northern Ireland where babies were born in June and July 2023, and where fieldwork started when the oldest babies were 10.5 months, babies were 8-14 months (mean age 10.5 months).

Most of the achieved sample babies were aged between 9-13 months old (70.7%). The babies whose parents took part in the online follow-up survey were older than those that took part in main fieldwork (between 13-19 months of age). Some babies in England and Wales are also older because of interviews that had to be redone because of a fraudulent interviewer (see section 3 for more details).

The distribution of the children's ages at interview is detailed in table 1.

Table 1. Distribution of child's/children's age at interview in months (cagemths – note child number is listed after 'c' in the parent dataset) of interviewed families by country of interview (dvctry_)

	England	Wales	Scotland	Northern Ireland	Total (UK wide)
9 months	112 (10.8%)	14 (4.9%)	0 (0%)	69 (21.4%)	195 (9.9%)
10 months	307 (29.6%)	77 (26.7%)	0 (0%)	135 (41.8%)	519 (26.3%)
11 months	254 (24.5%)	80 (27.8%)	4 (1.2%)	75 (23.2%)	413 (20.9%)
12 months	119 (11.5%)	48 (16.7%)	55 (16.9%)	16 (5.0%)	238 (12.1%)
13 months	72 (6.9%)	14 (4.9%)	116 (35.6%)	10 (3.1%)	212 (10.7%)
14 months	44 (4.2%)	13 (4.5%)	57 (17.5%)	18 (5.6%)	132 (6.7%)
15 months	53 (5.1%)	9 (3.1%)	39 (12.0%)	0 (0%)	101 (5.1%)
16 months	47 (4.5%)	20 (6.9%)	20 (6.1%)	0 (0%)	87 (4.4%)
17-19 months	31 (3.0%)	13 (4.5%)	35 (10.7%)	0 (0%)	79 (4.0%)

2.4 Fieldwork design

2.4.1 Types of informants

Three different types of interviews for up to four different respondents were developed for each family (or birth event):

1. **Primary Informant (PI)**: this was a 60-minute interview completed with an interviewer, primarily face-to-face, but alternative modes of interview included telephone, Teams and web in a small number of cases. The parent who completed this interview was living in the child's main household, and ideally was the parent who spent the most time looking after the child. This respondent was asked about their own characteristics, their household's characteristics and detailed questions about their child's characteristics. This

was the longest and most detailed interview type, and interviewers were instructed to try and always achieve this interview for each birth event.

2. **Own-household parent (OHP)**: this was a 40-minute interview completed with an interviewer (face-to-face, via telephone or via Teams) or online. This interview was for parents who did not live full-time or mainly in their child's main household. The questions asked covered the individual's characteristics, their household's characteristics and some questions about their relationship with their child.
3. **Additional Informant (AI)**: this interview was for a parent or partner who lived with the Primary Informant (**AIMain**) or a live-in partner of an Own-Household Parent, who had some contact with the child (**AIOHP**). Most often this was the baby's other parent, but they could also be a new partner (i.e. step-parent). This interview was 30-minutes long and could be completed with an interviewer face-to-face, via telephone, Teams, or online. The interview asked about the AI's own characteristics and their activities with the child.

The type of interview assigned to each participant was determined during a short screener survey when the interviewer first made contact with the parent at the household. The screener also confirmed eligibility for the study. The data from the screener is not included as part of the shared data.

This set-up meant that the study could: accommodate different types of family structures; make the interview structures efficient so that questions that only needed to be answered once were not answered multiple times within families; ensured that only parents with legal parental responsibility gave consents for their child; and allowed a flexible structure that could accommodate changes in household composition from the information from the sample frame and some parental choice about who was better placed to answer questions about their child. In particular, the PI interview was assigned on the basis of which parent spent most time with the child, rather than prioritising mothers over fathers.

The number and types of eligible informants depended on the household composition as reported in the screener interview. All main households had someone eligible to complete the PI interview, not all main households had someone eligible to complete the AIMain interview. All OHP households had someone eligible

to complete the OHP interview, not all OHP households had someone eligible to complete the AIOHP interview. Interviews were not obtained with all eligible respondents within a responding household due to non-response at parent interview level.

In practice, most often the PI interview was completed by the child's biological/birth mother (92% of PIs) and most often the AI interview was completed by the child's biological/birth father (59% of AIs). OHPs were nearly entirely biological/birth fathers (97% of OHPs).

Implications for data users

You do not need to sort the data by the type of interview a parent did – the data has already been sorted into parent A, B and C for you (_pa, _pb and _pc suffixes, see more details on this in chapter 5) – but we have included some further information below on how this informant design affected the data to help users understand why some parents have missing data on some variables.

Although up to four interviews were possible, there were no families with four respondents. There was one family where three interviews were obtained (PI, AIMain and OHP). For all other families, there are either one or two respondents (sometimes across two households).

The total number of parents per birth event (family) who did an interview is in variable 'total_parents_perbirth' and the interviews they did are listed together in 'dv_birth_event_parent_roles'. There is also a variable for the total number of parents per household who did an interview in 'total_parents_perhh' and the interviews that were done in the household is in 'dv_hh_parent_roles'.

The type of interview completed by an individual parent is listed in 'parent_role'. In 56 cases, the listed parent role (PI/AI/AIOHP/OHP) does not fully align with the questions completed by the parent according to the questionnaire. These discrepancies are due to parents being given the incorrect type of interview for their circumstances during fieldwork, and Ipsos later corrected the 'parent_role' to reflect the type of interview the parent should have done. The most common discrepancies you will see in the data are:

- Parents have an ‘AI’ parent role but have completed the detailed child questions which were only supposed to have been answered in the PI interview. This has arisen because sometimes interviewers accidentally gave these second parents a link to the PI version of the web questionnaire, when they should have been given the AI version.
- Parents have an ‘AI’ or ‘PI’ parent role but seem to have completed questions relating to the child’s second household which were only given in AIOHP or OHP interviews. This has arisen because the household was incorrectly assigned as the child’s second household by the interviewer during the screener.

Another thing to note is that there are a small number of families where there are more than one parent completing the same type of interview, specifically two parents each doing PI interviews (7 birth events/families). Two of these were the result of an error in how the online follow-up survey was administered (see section 2.4.3) resulting in two parents who should have done an AI interview (because the other parent had already completed the PI interview) completing the PI online follow-up survey instead. The other five are not the result of a fieldwork error but are because of how the screener allocated interviews: the screener uses the child’s residency to decide if a household was a ‘PI’ (i.e. child’s main household) or ‘OHP’ (i.e. not the child’s main household) household. However, there were instances where parents living at separate addresses both said their child lived with them most or all of the time. In these cases, both parents were given PI interviews.

More information about the questions included in each interview can be found in the questionnaire chapter (chapter 4), and more information about the parent data structure can be found in chapter 5.

2.4.2 Contact strategy

If a parent had not opted-out of the study after the first notification mailing, they were allocated to an interviewer who would post them an invitation mailing 3-4 days before they visited the household.

Interviewers were required to make a minimum of six face-to-face calls at all the sampled households they had been provided with at the start of fieldwork. In addition

to the requirement to make at least six calls to contact the named parents, if interviewers achieved an interview with the PI/OHP and there was still an Additional Informant (AI) interview needed at that household (or the AI was interviewed, but the PI/OHP had not), interviewers made additional visits or calls to interview the other parent in the household.

Where interviewers identified that the family had moved from the issued address, they carried out tracing in the field, which primarily consisted of asking current occupiers of the family's address (see the Generation New Era technical report).

Where only one of two named parents had moved, interviewers attempted to conduct the interview with the named parent who still lived there, and either asked that parent for the address of the other parent as part of the interview, or asked them to pass a forwarding letter on to them.

Interviewers were provided with an Additional Informant (AI) letter, which was used to help interviewers to obtain an interview with the AI, either in-person or online, and to provide the login details for the online survey. Although the preferred mode for interviewing the AI was face-to-face, they could also take part online to allow more flexibility.

2.4.3 Online follow-up survey

At the end of the main fieldwork period, all non-responding households (i.e., a household where neither a PI, AI nor OHP had taken part) received an invitation in the post to an online follow-up survey. Only one parent in the household was asked to complete the online follow-up survey, making their informant allocation either a PI or an OHP depending on whether they lived in their child's main household. In two cases, two parents who should have completed an AI interview completed the online follow-up survey as PIs because the survey was sent in error to some households where a PI had already taken part in main fieldwork. These are two of the seven 'two PI' families. More information on this can be found in the Generation New Era Technical Report.

The online follow-up survey for the Primary Informant was shorter than the main interview at 30-40 minutes long; the online follow-up survey for the OHP was

identical to the OHP's main interview. More information on the content of the questionnaire can be found in section 4.1.

The main implication of the online follow-up survey for data users is that there is more missing data for PIs and OHPs who completed the online follow-up version of the survey than PIs and OHPs who took part in main fieldwork.

2.4.4 Incentives

The study experimentally tested both 'unconditional' and 'conditional' incentives to identify which combination of incentives resulted in the highest response rates. Within each country, families were randomly allocated to three unconditional incentive conditions, and two conditional incentive conditions, for a total of six incentive groups. This means that all parents interviewed within the same family (birth event) received the same incentives. Unconditional incentives were sent once per household in the second 'invitation mailing' following the opt-out period. Around a third of the families were sent a £5 note per parent; another third were sent a baby's bib and the final third did not receive any unconditional incentive.

Conditional incentives were only given after a respondent had completed the interview/survey. Within each country, half of the families were offered a £10 voucher for each person who completed an interview, and the other half were offered a £20 voucher for each person who completed an interview. Face-to-face participants could choose to receive their voucher in the form of a physical gift card or digital 'e-voucher', and those completing the survey by telephone, Teams, or online received a digital 'e-voucher'.

The online follow-up survey had additional conditional incentives in some countries. In England and Wales, the conditional incentive remained the same as for the main fieldwork (either £10 or £20 conditional on completing the survey). This was because ethical approval was not received in time for a different design.

In Scotland and Northern Ireland, following the receipt of ethical approval, the conditional incentive was increased by £10 for everyone (on top of the original incentive condition, taking the conditional incentive to either £20 or £30). In addition, those living in more deprived areas (those in deciles 1-3 on the index of multiple

deprivation (IMD), where response rates had been lower), were offered an additional £20 on top of their original £10/£20 incentive condition (meaning that the incentive in these areas was £30 or £40).

This design aimed to provide insight into whether an increase in the conditional incentive offered for the online follow-up survey could improve response rates, and if the extra incentive for those in IMD 1-3 additionally improved their response rates.

3. Response

3.1 Overall response and outcome codes

In ELC-FS, a family (or birth event) was considered to have a 'productive' outcome for the purposes of response rate calculations if at least one interview in the child's main household (a PI or AIMain interview) had been achieved. **We call these families 'productive families' in this user guide and you can use variable 'dv_birth_event_mainhhinterview' to identify them**. An interview in the child's main household (a PI or AI interview) was achieved among 1918 families ('productive families', 1917 families in safeguarded data with triplet family excluded), and there were a further 15 families where only an OHP/OHP and AIOHP interview(s) were achieved. Therefore, while there are 1933 families in the data set, only 1918 of these are considered 'productive families' in calculations of response rates.

To calculate response rates using this number of productive families, we have two denominators 1) eligible birth events who were sent a notification mailing (3730 families) and 2) eligible birth events issued to fieldwork following opt-outs from this notification mailing (3633 families). 'Eligible' means some families from the original sample were excluded for reasons like the family had moved overseas, a parent/child had died, or because the family circumstances were too sensitive (e.g. because of adoptions).

These denominators also exclude another group of families: those affected by an interviewer who was found to have been falsifying parts of interviews. All interviews achieved by this interviewer were deleted, though some cases were reissued later to achieve a legitimate interview, and **these interviews appear in the data marked by the 'reinterviewed_cases' variable**. In total there were 86 reinterviews achieved (52 PIs, 32 AIs and 2 OHPs), across 53 birth events. The reinterviewed cases are included in the total number of interviewed families stated earlier (1933) but are not, however, counted as part of the total number of eligible families in Table 2 below because the entirety of the cases issued to the fraudulent interviewer needed to be excluded in the calculation of the study/survey response rate.

You do not need to treat these reinterviewed cases separately or differently in your analysis, unless you are concerned about the impact the reinterview might have had on the quality of the survey data (e.g. the implications of the same person answering the same question again).

Excluding the 53 reinterviewed productive families because of the fraudulent interviewer, and another 22 productive families that were in part worked by the fraudulent interviewer but were ultimately legitimate, **the total number of productive families used in response rate calculations is 1843 families**. 1843 families from the initial eligible sample sent a notification mailing of 3730 gives a study response rate of 49%. 1843 families from the eligible issued sample (post opt-out and exclusions) of 3633 gives a survey response rate of 51%.

Table 2 below shows the overall survey response for the families issued to fieldwork. It also shows that the main reason for non-response in main fieldwork (i.e. prior to online follow-up survey) was refusal (31%) followed by non-contact (14%).

Table 2. Overall family-level response for baby's main household issued sample

Main household defined as household where baby lives all/most of the time; where unknown, Address 1 (mother address from sample) used as baby's main household.

		Country Breakdown			
	Total	**England**	**Wales**	**Scotland**	**Northern Ireland**
Total eligible families issued to fieldwork	3633	2064	447	539	583
Productive families	1843	958	246	319	320
Survey response rate	51%	46%	55%	59%	55%
Main fieldwork outcomes and rates					
Refusals	1119	626	143	159	191
Refusal rate	*31%*	*30%*	*32%*	*29%*	*33%*
Untraced/not located	306	229	29	29	19
Untraced/not located rate	*8%*	*11%*	*6%*	*5%*	*3%*
Non-contact/other unproductive	491	309	44	60	78
Non-contact rate	14%	15%	10%	11%	13%
Note: Productive families are those eligible families with at least one productive PI or AI interview in the main household. Online follow-up non-respondents are classified as 'non-contact' in the main fieldwork section of the table. All cases allocated to the fraudulent interview are excluded from this table, including those who were successfully reinterviewed later or were ultimately found to be legitimate.					

Main fieldwork stage

Among the 3633 eligible issued families, there were 1717 productive families who took part in main fieldwork (excluding re-interviews due to the fraudulent interviewer) i.e. a full or partial interview with a PI or an AI in the child's main household. By the end of main fieldwork, the survey response rate was 47% (44% in England, 52% in Wales, 54% in Scotland and 51% in Northern Ireland).

Online follow-up stage

The online follow-up survey was issued to any household (PI or OHP household) where no interview had been achieved. 1324 households were issued in the online follow-up phase (787 England, 212 Wales, 179 Scotland, 146 Northern Ireland). 128 additional families completed an interview in the online follow-up (127 PIs and 1 OHP; 60 families in England, 15 Wales, 28 Scotland, 25 Northern Ireland). 128 responding families of 1324 households issued gave a response rate for the online follow-up survey of 10% (8% England, 7% Wales, 16% Scotland, 17% Northern Ireland). The online follow-up stage increased the overall survey response rate to 51%, a 4 percentage points increase (to 46% England, 55% in Wales, 59% in Scotland and 55% Northern Ireland). Please note there were two additional interviews achieved in the online follow-up stage (taking total interviews to 130) in households where a parent had already taken part during main fieldwork due to an error in administration. These parents are therefore not included in the response rate calculations for the online follow-up stage. More detail on this can be read in section 2.4.1 and the Technical Report.

3.2. Mode of completion

Including all cases (not only productive families, and also including reinterviewed cases/cases that were legitimate associated with the fraudulent interviewer, the triplet family, and main and online follow-up fieldwork) in-person, face-to-face interviewing (CAPI) was the most common mode across all types of informant interviews. This mode was most common for PIs (86% of interviews) compared to AIs in a PI household (58%) and OHPs (63%). Web (CAWI) during main fieldwork was much more common among AIs in a PI household than other interviews because interviewers were encouraged to offer this option to AIs if they felt this was the most likely way to secure their interview.

Survey mode is denoted by the int_mode variable in the survey data set. Fieldwork stage (i.e. main fieldwork or online follow-up fieldwork) is denoted by the variable fieldworkstage.

Table 3. Mode of response (int_mode) by parent interview role (parent_role) for all parents

	In-person	Telephone	Teams	Web during main fieldwork	Web during online follow-up fieldwork	Total
PI	86%	6%	1%	<1%	7%	1910
AI in the main/PI household	58%	8%	1%	34%	No interview	1,152
OHP	63%	17%	2%	17%	2%	60
AI in OHP household	100%	No interview	No interview	No interview	No interview	4

4. Questionnaire

4.1 Overview of questionnaire topics, administration and length

The ELC-FS questionnaire was developed following an extensive scientific consultation with different stakeholders in 2021. More information about the decisions taken for the questionnaire following consultation can be read here on the ELC-FS CLS website.

The ELC-FS questionnaire was comprised of different modules which covered household relationships, housing, parent's background (education, employment ethnicity, health), information about other parents or partners, household income, self-completion for sensitive topics (computer-assisted self-interviews - CASI), childcare, child health, the parent-child relationship and contact information. Consents for data linkage were also collected for all participants and their children, and consent to give a saliva sample was asked to some parents (neither of these are included in the data deposit).

Broadly speaking, the AI questionnaire only included modules with questions about the parent themselves, the OHP questionnaire included modules about the parent and their household, and the PI received all modules meaning they were the only informant who answered detailed questions on the child. Sometimes questions needed to be answered by a specific parent (e.g. questions about pregnancy history were only asked to the child's biological mother). These kinds of questions appeared in the different informant interviews, but routing was used to ensure that the correct parent answered the questions regardless of which interview they did. A breakdown of which questions were asked to each informant is in Table 4.

An interview was classified as a 'complete' interview if the respondent reached the end of the questionnaire. An interview was classified as 'partial' if the respondent completed at least the first two modules of their interview but did not reach the end: beyond the household grid module for PIs and OHPs and beyond the background module for AIs. Anyone who completed fewer modules than this was classified as

unproductive and has not been included in this data. **Users can see whether an interview was 'full' or 'partial' using the variable 'intstatus'.**

Interviews could be done in languages other than English with the help of a translator. Respondents could request an interviewer who spoke their language, use someone they knew to translate/support the interview, or use Ipsos's relay service. Showcards were also translated into 10 different languages, accompanying translations done for all recruitment materials into these same ten languages. **Variable 'oetr' indicates whether the interview was conducted in a language other than English (76 parents' interviews (3%)). Variable 'oewh' can be used to establish who the translator was, and 'oets' can be used to established whether translated materials were used as part of the interview. The language of the translated materials ('oetm') and the language the interview was conducted in ('oeco') is available in the secure access deposit.**

The computer assisted self-interview (CASI) module consisted of relatively more sensitive or personal questions, including parent-infant bonding, deprivation measures, pregnancy history, adverse life events, social support, loneliness, satisfaction with their romantic relationship, mental health, smoking, drinking and vaping. Those who completed an interviewer administered survey (i.e., CAPI, CATI, Teams) were told that the next set of questions were considered more sensitive. If the interview was face-to-face, they were instructed to respond using a tablet that was given to them. If they required assistance by someone else (e.g., a family member for translation), routing was used to remove the most sensitive questions (e.g., about the couple relationship if the person assisting was the person's partner). If the person completed their interview over the telephone, showcards were used which were sent to the participant in advance. If the participant could not locate or use their showcards, then the interviewer read the question and response options over the phone. Some informants refused to participate in this module altogether, as there was an option to spontaneously refuse the CASI module in face-to-face, Teams and telephone interviews.

Informants who took part during the online follow-up phase completed an abbreviated version of the questionnaire. The abbreviated questionnaire was designed to include the core measures which would be used most widely by

researchers. An initial question established whether the parent lived in the child's main household or not, determining whether the parent responding would be classified as a PI or an OHP. While the online-follow up survey was broadly the same for both types of informant, routing was used so that OHPs were not asked questions about their child's details/consents, which were only asked to PI informants as in the main survey. Table 4 shows which content was included in the abbreviated online follow-up questionnaire.

Users can see which cases completed the abbreviated version of the questionnaire via the 'fieldworkstage' variable.

Table 4. Questionnaire content

Module	Topics covered	Primary Informant	Own Household Parent	Additional Informant	Online follow-up survey
Introduction	**Interview set up**	**x**	**x**	**x**	**x**
	Informant assignment/checks	**x**	**x**	**x**	**x**
	Consent to take part Includes information about data linkage for those in opt-out group	**x**	**x**	**x**	**x**
	Informant details Name, address, date of birth, age, gender, sex at birth,	**x**	**x**	**x**	**x**
	Baby details and relationship Name, date of birth, sex, relationship to child	**x**	**x**	**x**	**x**
	Other parent's name			**x**	
Household grid	**List of other household members** Name, date of birth, age, gender, relationship to cohort child, relationship to respondent	**x**	**x**		**x**
	Parent and partner details	**x**	**x**		**x**

Module	Topics covered	Primary Informant	Own Household Parent	Additional Informant	Online follow-up survey
	Identification of other biological parent in household and identification of partner in the household (whether biological parent or not). For each (if different): sex at birth, whether resident in another household, whether child stays with them in other household, how often partner/parent stays in respondent's house, current relationship status, length of cohabitation, whether currently or ever married. For parent who isn't a partner: whether ever in a relationship, length of relationship, date of separation. For children with no biological parents in household: whether they have any contact with biological parents				
	Details of any parents not in household list Whether any contact/involvement with parent not on household list (PI interview only), name, gender, sex at birth, date of birth, age, whether respondent in a relationship with this parent, whether ever in a relationship, length of relationship, whether currently or ever married.	**x**	**x**		**x**
	Contact between child and own household parent	**x**	**x**		**x**

Module	Topics covered	Primary Informant	Own Household Parent	Additional Informant	Online follow-up survey
	How often OHP sees child and for how long, how often baby stays overnight with OHP, how often OHP stays overnight in child's main household				
	Non-resident children Names, gender, date of birth, age, relationship to cohort child, relationship to respondent	**x**	**x**		**x**
Housing	**Housing** Housing tenure, number of rooms, access to a car and outdoor space	**x**	**x**		
	Languages spoken at home	**x**	**x**		
Background	**Employment** Main activity, whether ever employed, whether self-employed, usual hours worked a week, full/part time, whether out of work due to health, when last/current job started, when last job ended, job title, what mainly makes or does, qualifications needed for job, how many supervisees, whether any managerial role, how many employees at business, SOC2020 and SIC codes	**x**	**x**	**x**	**x**

Module	Topics covered	Primary Informant	Own Household Parent	Additional Informant	Online follow-up survey
	Parental leave Whether and when stopped working before baby's arrival, whether took or currently on parental leave, when leave finished/will finish, whether planning to return to work, how job has changed if already returned	x	x	x	x
	Education Age left full time education, academic and vocational qualifications	x	x	x	x
	Ethnicity (own, child) Country of birth, own ethnicity, child's ethnicity (PI only)	x	x	x	x
	Religion	x	x	x	
About Partner	**Employment** Main current activity, working full/part time, whether out of work due to poor health	x	x		x
	Parental leave	x	x		

Module	Topics covered	Primary Informant	Own Household Parent	Additional Informant	Online follow-up survey
	Whether took/currently on parental leave, when parental leave finished/will finish, length of parental leave, whether plans to return to work				
	Highest qualification	**x**	**x**		**x**
	Ethnicity	**x**	**x**		
	Health Whether any long-term health conditions and impact on daily life	**x**	**x**		
Family income	**Benefits** Receiving universal credit, which benefits receiving	**x**	**x**		**x**
	Family income	**x**	**x**		**x**
	Keeping up with bills in last six months	**x**	**x**		**x**
Child's health	**Fertility treatments** Type of fertility treatment, time to conception	**x**			
	Birth & delivery	**x**			**x**

Module	Topics covered	Primary Informant	Own Household Parent	Additional Informant	Online follow-up survey
	Due date, whether born early/late, whether cared for in neonatal unit, when came home from hospital				
	Birth weight	**x**			**x**
	Health General health, A&E visits, longstanding/developmental conditions, hospital visits for condition	**x**			**x**
	Diet Breastfeeding and solids introduction	**x**			
	Sleep Where usually sleeps, how often wakes up in night, length of sleep, whether child's sleep a problem	**x**			
	Extent of screen use and crying	**x**			**x**
	Development Smiling, sitting, standing, putting hands together, picking up objects, passes toy back and forth, take a few steps, moves from one place to another	**x**			

Module	Topics covered	Primary Informant	Own Household Parent	Additional Informant	Online follow-up survey
Child activities & temperament	**Parenting engagement** How often looks after child alone, changes nappies, feeds them, soothes them, gets up during the night for child	**x**	**x**	**x**	**x**
	Temperament Short form Infant Behaviour Questionnaire	**x**	**x**	**x**	
	Play activities Active physical play, gentle physical play, toys, pretend games, turn-taking games, showing pictures, reading, noisy play, singing, going outside, cuddles, talks to child	**x**	**x**	**x**	**x**
CASI	**Pregnancy history (biological birth mother only)** Whether currently pregnant, how many previous pregnancies, how many resulted in a live birth	**(x)**	**(x)**	**(x)**	
	Parent-infant bonding & parental stress	**x**	**x**	**x**	**x**
	Disadvantage indicators Housing conditions, how doing financially, whether skips meals, unable to afford essential baby items	**X**	**x**	**x**	**x**

Module	Topics covered	Primary Informant	Own Household Parent	Additional Informant	Online follow-up survey
	Health General health, longstanding health conditions and impact on daily life,	X	x	x	x
	Life events during and after pregnancy	X	x	x	
	Social support, life satisfaction and loneliness	x	x	x	x
	Couple relationship Satisfaction and extent of disagreements about parenting	x	x	x	x
	Mental health PHQ-4 and Kessler-6, whether consulted a doctor or received treatment	x	x	x	x
	Smoking, alcohol, vaping During pregnancy, currently, whether ever smoked, whether smoked vaped in same room as baby, how many people smoke/vape in household near baby	x	x	x	x
	Info: sources of support	x	x	x	x
Childcare	**Childcare providers**	x			x

Module	Topics covered	Primary Informant	Own Household Parent	Additional Informant	Online follow-up survey
	Childcare type, when started and ended, time spent in childcare				
	Grandparent financial support	**x**			
	Service use Types of professionals and services used, how often, whether had 6-week health visitor review, problems accessing health visitors/GP when needed	**x**			**x**
Data linkage (opt in)	**Data Linkage-OPT IN group** Individual consents for health, education and social care linkages. Only PI gives child's consents.	**x**	**x**	**x**	**x**
Contact details and end	**Contact details**	**x**	**x**	**x**	**x**
	Contact details for non-resident parent	**x**			**x**
	Stable Contact	**x**	**x**	**x**	**x**
	Details of new address if moving	**x**	**x**		**x**
	Incentive and Outro	**x**	**x**	**x**	**x**

The PI main fieldwork interview (i.e. not online follow-up survey, excluding partial completes and outliers under 10 minutes/over 200 minutes) took 60 minutes to complete face-to-face or by telephone on average (n=1622). The survey was on average 10 minutes longer when done via Teams (n=110) and 10 minutes shorter when done via the web (n=4).

The AI interview (again excluding partial completes and outliers) took on average 30 minutes when completed face-to-face (n=667), by telephone (n=90) or online (n=379). It took 43 minutes on average when done by Teams (n=10).

The OHP interview (excluding partial completes and outliers) took on average 40 minutes to complete face-to-face (n=38). It was slightly shorter by telephone (36 minutes, n=9) and slightly longer by web (47 minutes, n=9) though numbers of respondents for these modes were small. There were no OHP Teams interviews.

The average completion time for the online follow-up survey was approximately 35 minutes for PI respondents. A single OHP respondent completed the online follow-up survey, which took 71 minutes.

Variations to questions across modes were modest and mostly limited to variations in the interviewer instructions about using showcards or what to read out, and variations in question wording to assist web self-completion. These differences can be seen in the questionnaire document accompanying this user guide.

Section 7 of this User Guide discusses how researchers should take mode effects into account in analysis.

The questionnaire was scripted and implemented by Ipsos. It was extensively tested both by Ipsos and CLS.

4.2 Other special features of the questionnaire

4.2.1 Occupation coding

Participants were asked to provide details about their current job, or the last job they had if not currently working.

All occupations were coded to the four-digit Standard Occupation Coding frame (SOC 2020). To minimise disclosure risk, 3-digit SOC codes are available as safeguarded data (End User Licence), and these are sometimes recoded as 2-digit codes where there are low counts. The 4-digit SOC codes are available under Secure Access (see Sections 5.1 and 5.14).

The National Statistics Socio-Classification, rebased on SOC2020 (NS-SEC) has also been derived from the SOC code. NS-SEC data is also presented in five, seven, eight and 13 analytic classes.

The UK Standard Industrial Classification of Economic Activities (SIC) has also been derived and is available under Secure Access, and a truncated version to three characters (or two where there are low counts) is available in the safeguarded data (EUL). Social grade has also been coded (A, B, C1, C2, D, E).

The additional office-based codes are for research via application to the CLS Data Access Committee: cls.ucl.ac.uk/data-access-training/data-access/accessing-data-directly-from-cls/.

The below table summarises occupation coding variables based on respondent's current job (or last job if not currently working) in the safeguarded data (EUL):

Description	**Variable name**
3-digit SOC code (per SOC2020)	soc2020_tr
Six category social grade	socialgrading
NS-SEC (SOC 2020)	dvnssec
NS-SEC 13 (SOC 2020)	dvnssec13
NS-SEC 8 (SOC 2020)	dvnssec8
NS-SEC 7 (SOC 2020)	dvnssec7
NS-SEC 5 (SOC 2020)	dvnssec5
3-digit SIC code	sic_tr

4.2.2 Geographical indicators

The full set of Early Life Cohort Feasibility Study geographical identifiers will be available from the UK Data Service through Secure Access.

A subset of these indicators is available within the Parent and Child level Safeguarded datasets (see Section 5 for information on available/future datasets). See Appendix 1 (A1.1) for a list of Safeguarded geographical variables.

4.2.3. Alignment with other 9-month birth cohorts

The ELC-FS questionnaire has been heavily informed by other 9-month surveys to allow for comparisons to be made between the datasets. In particular, the ELC-FS questionnaire encapsulates the entirety of the Children of the 2020s wave 1 survey (England only, fieldwork 2022). Users will also find many questions from the first wave of the UK-wide Millenium Cohort Study (fieldwork 2001) in the ELC-FS questionnaire to allow for cross-cohort comparisons to be made. A report on the similarities and differences between ELC-FS and other 9-month longitudinal surveys is forthcoming from CLS.

4.3 Scales and standardised measures

ELC-FS included several established standardised measures which are listed below. Where scales are used, scores for each scale have been derived and included within the data and are covered in this section. Further details regarding the derivation of the scores can be found in Appendix 1, and original wording used in the scales can be found in the ELC-FS Questionnaire.

4.3.1 Background and About Partner modules: Ethnicity

Ethnic group (ONS, 2021)

Ethnicity is measured in ELC-FS for each responding parent (PI, AI or OHP **pethnic_a**), for the cohort child in the PI interview (**ethnic2_a**) and reported for a co-resident partner by the PI or OHP in their interview (**ethnicp_a**). The ethnicity classification used in each is the ONS classification used in the 2021 census. The response options given to respondents were:

1. White - English/ Welsh/ Scottish/ Northern Irish/ British
2. White - Irish
3. White - Gypsy or Irish Traveller
4. White - Roma
5. Any other White background (Please write-in) [TEXTBOX: ETHNICw]
6. Mixed/multiple ethnic groups - White and Black Caribbean
7. Mixed/multiple ethnic groups - White and Black African
8. Mixed/multiple ethnic groups - White and Asian
9. Any other mixed/multiple ethnic background (Please write-in) [TEXTBOX: ETHNICm]
10. Asian/Asian British - Indian
11. Asian/Asian British - Pakistani
12. Asian/Asian British - Bangladeshi
13. Asian/Asian British - Chinese
14. Any other Asian background (Please write-in) [TEXTBOX: ETHNICa]
15. Black/African/Caribbean/Black British - African
16. Black/African/Caribbean/Black British - Caribbean
17. Any other Black/African/Caribbean background (Please write-in) [TEXTBOX: ETHNICb]
18. Other ethnic group - Arab
19. Any other ethnic group (Please write-in) [TEXTBOX: ETHNICo]

Pethnic_a, ethcni2_a and ethnic_a are only available in the Secure Access data. The ethnicity data in the safeguarded (End User License) dataset consists of derived variables using these three ethnicity variables to amalgamate ethnicity into different categories (4 categories (respondents only), 6 categories, 8 categories and 11 categories) for the respondent, the cohort child and for co-resident partners.

Variable name	**Variable label**
pethnic_a	(A) Respondent's ethnic group
ethnic2_a	(A) Cohort child's ethnic group
ethnicp_a	(A) Partner's ethnic group
dvcmethnic6	DV Cohort Member Ethnic Group - 6 category Census class

Variable name	Variable label
dvcmethnic8	DV Cohort Member Ethnic Group - 8 category Census class
dvcmethnic11	DV Cohort Member Ethnic Group - 11 category Census class
dvethnic4	DV Respondent ethnicity – 4 categories
dvethnic6	DV Respondent ethnicity – 6 category Census class
dvethnic8	DV Respondent ethnicity – 8 category Census class
dvethnic11	DV Respondent ethnicity – 11 category Census class
dvethnicp4	DV Partner's ethnicity – 4 categories
dvethnicp6	DV Partner's ethnicity – 6 category Census class
dvethnicp8	DV Partner's ethnicity – 8 category Census class
dvethnicp11	DV Partner's ethnicity – 11 category Census class

4.3.2 Child health module: Child's sleep

Sleep Habits Questionnaire (Seifer et al., 1996)

The Sleep Habits Questionnaire measures the sleeping habits of infants and young children. While the full questionnaire has 12 items, the ELC-FS questionnaire used two items:

1) How often the child wakes at night (**sleep4**)
   RANGE: 0..50 times per night
2) Child's total amount of sleep in hours and minutes (**sleep6**).
   RANGE: 0…24 hours
   RANGE: 0…59 minutes

Only PIs completed these questions. The questions were completed once for every cohort child.

Variable name	Variable label
sleep4	How often cohort child woke up during night in the past 2 weeks

Variable name	Variable label
sleep6hr	How much time cohort child sleeps at night- hours
sleep6min	How much time cohort child sleeps at night- minutes

Brief Infant Sleep Questionnaire, BISQ (Sadeh, 2004)

The BISQ measures the sleeping habits of children aged 0-36 months. The ELC-FS questionnaire used two items:

1) Where the child usually sleeps (**sleep1_a**)
    1. Infant bed/cot in their own room
    2. Infant bed/cot in parents' room
    3. In parents' bed
    4. Infant bed/cot in room with sibling
    5. Other (please write-in) [TEXTBOX: SLEEP1O]

2) Whether the parent considers their child's sleep to be a problem (**sleep8a**)
    1. Not a problem at all
    2. A very small problem
    3. A small problem
    4. A moderate problem
    5. A serious problem

Only PIs completed these questions. These questions were asked once for each cohort child. sleep1 was a multicoded variable meaning more than one response option could be chosen.

Variable name	Variable label
sleep1_a	Where cohort child usually sleeps
sleep8a	How serious a problem is cohort child's sleep

4.3.3 Child activities & temperament module: Parent-child play and interactions

Bronte-Tinkew measure of father involvement (Bronte-Tinkew et al., 2008)

The Bronte-Tinkew measure of father's involvement assesses how often fathers contribute to 19 different child-related activities. The measure was included in the 9-month wave of the US ECLS-B study. We included 1 item from the measure about how often fathers soothe their child when upset (**MCCTe/OHPMCCTe**), as many other components of the scale were covered in other measures included in this module. This question was only asked in AI or OHP interviews, as the PI interview was preferentially given to the child's main caregiver. The response options for this question were:

mccte:

1. More than once a day
2. Once a day
3. A few times a week
4. Once or twice a week
5. Less than once a week
6. Never

ohpmccte:

1. More than once a day
2. Once a day
4. Less often than once a day
6. Never

mccte is the AI version of the question, and ohpmccte is the OHP version of the question. The question wording differed slightly between the two, where AIs were prompted to think about a typical week, whereas OHPs were prompted to think about the days when they spend time with their baby. This is why the response options differ slightly between the two.

Variable name	Variable label
mccte / ohpmccte	How often soothe cohort child(ren)

The Parent Play Questionnaire (PPQ) parent-child play subscale (Ahmadzadeh et al., 2020)

The PPQ is a parent-reported measure of play on three sub-scales: frequency of parent-child play, frequency of child's digital media use and parental attitudes towards play. The ELC-FS included all items from the frequency of parent-child play subscale in each informant's interview (PI, AI and OHP, though the latter had some modification to response options depending on how often the OHP saw their child).The items include frequency of active physical play (**playf1, ohpplayf1**), gentle physical play (**playf2, ohpplayf2**), playing with toys (**playf3, ohpplayf3**), pretend games (**playf4, ohpplayf4**), turn-taking play (**playf5, ohpplayf5**), play with books (which the study team split into two items to align with ALSPAC study – showing pictures in books **playf6, ohpplayf6** and reading stories **playf7, ohpplayf7**), noisy play (**playf8, ohpplayf8**) and singing (**playf9, ohpplayf9**). The PLAYF version asked to PIs and AIs was asked over a timeframe of a typical week, as in the PPQ scale. The OHPPLAYF version asked to OHPs and AIOHPs was asked in reference to a typical day spent with their baby, because the time spent with the child is more variable for this group.

The response options for each item were:

playf versions:

1. Never
2. Less than once a week
3. Once or twice a week
4. Several times a week
5. Once or twice a day
6. Several times a day

ohpplayf versions:

1. Never
2. Less often than once or twice a day
3. Once or twice a day
4. Several times a day

In addition to the items in the PPQ scale, the study team added three additional types of play to align with how the questionnaire module on parent-child play was asked in the first wave of the children of the 2020s study (**playf/ohpplayf10-playf/ohpplayf12**). These included how often parents took their child outside, cuddled them and spoke to them. The same response options were used for these items. These items are included in the table below for completeness but please not they are not part of the PPQ scale.

A derived variable (**dvhomelearnscore**) is available. This derived variable represents a combined home learning score, calculated by summing frequencies with which parents reported doing each of the home learning activities they were asked about. To calculate this derived variable, frequency response options were first transformed into numeric scores (“Never” = 1, "Several times a day" = 6) and then summed across the 12 activities (playf1 to playf12). Higher combined home learning scores indicate a higher frequency of home learning activities. Please note that this derived score contains one extra item relative to the same score in the first wave of the Children of 2020s study (only one item was asked about physical play rather than two). Only PLAYF responses were used for this score.

Variable name	**Variable label**
playf1/ohpplayf1	How often engages in active physical play with cohort child(ren)
playf2/ohpplayf2	How often engages in gentle physical play with cohort child(ren)
playf3/ohpplayf3	How often plays with toys with cohort child(ren)
playf4/ohpplayf4	How often plays pretend games with cohort child(ren)
playf5/ohpplayf5	How often does turn-taking play with cohort child(ren)

Variable name	Variable label
playf6/ohpplayf6	How often shows pictures in books to cohort child(ren)
playf7/ohpplayf7	How often reads cohort child(ren) stories
playf8/ohpplayf8	How often engages in noisy play with cohort child(ren)
playf9/ohpplayf9	How often sings to cohort child(ren)
playf10/ohpplayf10	How often takes cohort child(ren) outside
playf11/ohpplayf11	How often cuddles with cohort child(ren)
playf12/ohpplayf12	How often talks to cohort child(ren) about activities
dvhomelearnscore	DV Combined home learning score

4.3.4 Child activities & temperament module: Child temperament

Infant Behavior Questionnaire (IBQ)

The Infant Behavior Questionnaire is designed as a measure of temperament for children aged 3-12 months. The shortest version of the IBQ is 37-items across 3 broad scales – Surgency, Negative Emotionality, Effortful control (Putnam et al., 2014). The Children of 2020s study used a combination of item-response theory, reliability assessments and exploratory factor analysis to further condense these items. They used data collected on the very short IBQ questionnaire (37-items) by the APrON study for this exercise – a Canadian prospective longitudinal cohort study on gestational nutrition and development which measured IBQ between 6-12 months using a combination of the short (91-items) and the very short (37-items) IBQ questionnaires. Following validation exercises, a selection of 14 items was derived, which are included in the ELC-FS questionnaire (**ibq1-ibq14**). These questions were asked to all parents (PI, AI and OHP interviews) as long as they progressed through a check that they had seen their child in the last week (**ibqsee**). Each question was asked once per cohort child.

More information on this analysis is available from the Children of 2020s study team on request.

The response options for each item were:

1. Never (I saw my baby in this situation in the last week, but they never responded this way)
2. Very Rarely
3. Less Than Half the Time
4. About Half the Time
5. More Than Half the Time
6. Almost Always
7. Always
8. Does Not Apply (I did not see my baby in this situation in the last week)

A derived variable (**dv_ibq_cm1 – dv_ibq_cm3**) which contains the *average* score for each cohort child on the Infant Behaviour Questionnaire scale is available (see Appendix 1 for logic).

Variable name	**Variable label**
ibq1	How often cohort child laughed when tossed around playfully
ibq2	How often cohort child clung to parent when introduced to unfamiliar adult
ibq3	How often cohort child played with 1 toy for 5-10 minutes
ibq4	How often cohort child laughed when put into bath
ibq5	How often cohort child fussed when time for bed
ibq6	How often cohort child cried when waking up from sleep
ibq7	How often cohort child laughed during peekaboo
ibq8	How often cohort child was angry when left in crib
ibq9	How often cohort child looked at pictures in books/magazines for 5 minutes or more
ibq10	How often cohort child smiled/laughed when given a toy

Variable name	Variable label
ibq11	How often vocalised when hair was being washed
ibq12	How often cohort child refused to go to unfamiliar person
ibq13	How often cohort child cried when unable to get your attention
ibq14	How often cohort child stared at mobile/picture in crib for 5 minutes or longer
dv_ibq_cm1-dv_ibq_cm3	DV CM Average score of Infant Behaviour Questionnaire scale

4.3.5 CASI and 'About partner' modules: Parental disability

ONS long lasting health conditions and illnesses: Impairments and Disability (ONS, 2021)

ELC-FS included a sub-set of the ONS harmonised set of questions on Long-lasting Health Conditions and Illnesses including Impairments and Disability. This question was asked in each respondent's questionnaire (PI, AI, OHP) and asked to PIs and OHPs about their co-resident partner.

loil and lolm items listed below are used to derive variables indicating whether the respondent is disabled or notusing the Equality Act 2010 definition (**dvdisabilityea**). According to the Equality Act 2010 definition, a cohort member is considered to be disabled if they report a longstanding illness (loil) and have a reduced ability to carry out day-to-day activities as a result of their illness (lolm).

Variable name	Variable label
loil	Any physical or mental health conditions or illnesses lasting or expected to last 12 months
lolm	Whether illnesses/conditions reduce ability to carry out day to day activities
pgh1	Whether partner has any long-term health conditions
pgh2	Whether partner's health conditions reduce ability to do day-to-day activities

Variable name	Variable label
dvdisabilityea	DV: Disability classification Equality act (2010)

4.3.6 CASI module: Parental mental health

Generalised Anxiety Disorder 2-item, GAD2 (Kroenke et al. 2007)

The Generalized Anxiety Disorder 2-item (GAD-2) is a brief initial screening tool for generalized anxiety disorder. These items were asked in every parent questionnaire (PI, AI, OHP).

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

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

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

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

Variable name	Variable label
gad2phq2a	Whether nervous, anxious or on edge over last 2 weeks
gad2phq2b	Whether not being able to stop or control Worrying in the last two weeks
dvgad2	DV Generalised Anxiety Disorder 2-item

Patient Health Questionnaire 2-item, PHQ-2 (Kroenke et al, 2003)

The PHQ-2 enquires about the frequency of depressed mood and anhedonia over the past two weeks. The PHQ-2 includes the first two items of the PHQ-9. These items were asked in every parent questionnaire (PI, AI, OHP).

Respondents are asked how often they have been bothered by problems over the last 2 weeks: c) “Little interest or pleasure in doing things”; and d) “feeling down, depressed or hopeless”, with the following response options:

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

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

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

Variable name	**Variable label**
gad2phq2c	Whether had little interest or pleasure in doing things in the last 2 weeks
gad2phq2d	Whether feeling down, depressed or hopeless in the last 2 weeks
dvphq2	DV Patient Health Questionnaire 2-item

Kessler Psychological Distress Scale K6 (Kessler et al., 2002)

The Kessler Psychological Distress Scale (K6) is a brief screening scale that measures non-specific psychological distress. K6 is a 6-item short form of a longer 10-item scale (K10). These items were asked in every parent questionnaire (PI, AI, OHP).

The scale is scored on a 5-point scale for each of the six items. The score for each response option is:

0 = None of the time
1 = A little of the time
2 = Some of the time
3 = Most of the time
4 = All of the time

This gives a total score range from 0 to 24 (**dvkessler**). A score of 13 or higher is often used as a cutoff to indicate a probable serious mental illness.

Variable name	Variable label
kes1	How often felt nervous
kes2	How often felt hopeless
kes3	How often felt restless or fidgety
kes4	How often felt so depressed nothing could cheer up
kes5	How often felt everything was an effort
kes6	How often felt worthless
dvkessler	DV Kessler K6 Mental Health scales

4.3.7 CASI module: Loneliness and social support

Loneliness direct measure (ONS, 2018)

The ELC-FS uses the ONS-recommended single-item direct measure of loneliness which has been used in the Community Life Survey. These items were asked in every parent questionnaire (PI, AI, OHP). The question asks the respondent how often they feel lonely with possible response options:

1. Often or always
2. Some of the time
3. Occasionally
4. Hardly ever
5. Never

Variable name	Variable label
lonely	Feeling lonely

Brief form of the Perceived Social Support Questionnaire, F-SozU K-6 (Kleim et al., 2014)

The questionnaire measures perceived social support by shortening a well-established German questionnaire (F-SozU K-14) of 14 items into 6. The six items are ranked on a scale from 1-5, and the score across the six can be summed to provide a total estimate of perceived social support.

The six items were asked in each parent's interview (PI, AI and OHP).

Variable name	Variable label
ssq1	Support from others
ssq2	Have someone to count on
ssq3	Can borrow from someone
ssq4	Have people they enjoy doing things with
ssq5	Ask friends/family for help when sick
ssq6	Know who to go to when feeling down
dvssq	DV Perceived Social Support Questionnaire (FSozUK-6)

4.3.8 CASI module: Life events

List of threatening experiences (Brugha et al., 1985)

Brugha et al. developed the list of threatening experiences following an assessment of 67 possible life events, which found that 12 of these accounted for 77% of all life events rated as having a marked or moderate long-term threat. These items were included in the ELC-FS PI questionnaire, except for the final item about whether something valuable was lost or stolen in place of a different question about experiencing serious housing difficulties or being made homeless (**liev10a/liev10b**) which had been an important measure in the CLS COVID-19 surveys. Other

adjustments were also made to better align the question to modern contexts, align with other CLS surveys and gain better temporal precision:

1) In the original scale there was one item about whether the respondent's marriage had dissolved and another about whether subject broke off a steady relationship but these were combined (**liev5a/liev5b**).
2) In the original scale there was one item about whether the respondent had been unemployed for more than a month and one item about whether the respondent had been sacked. The former question was dropped as ELC-FS has other questions about being unemployed and work around the time of the baby's birth, and the latter was adapted to align with the question asked in the Covid Social Mobility and Opportunities Survey to ask both about losing job or losing a business (**liev7a/liev7b**).
3) Each question item was repeated to better pinpoint when exactly the threatening experience occurred. First the questions were asked in relation to the period of pregnancy with the cohort child (all items ending in A), and then they were repeated in relation to the period after the child's birth (all items ending in B).

Variable name	**Variable label**
liev1a/liev1b	Serious illness/injury to self
liev2a/liev2b	Serious illness/injury to close relative
liev3a/liev3b	Death of child, parent or partner
liev4a/liev4b	Death of other relative or friend
liev5a/liev5b	Separation/break-up/divorce from partner (adapted version of original item)
liev6a/liev6b	Serious problem with friend, neighbour, relative
liev7a/liev7b	Lost job/business (adapted version of original item)
liev8a/liev8b	Major financial difficulties

Variable name	Variable label
liev9a/liev9b	Trouble with police/court

4.3.9 CASI module: Satisfaction with couple relationship

Couples Satisfaction Index, CSI-4 (Funk & Rogge, 2007)

The Couples Satisfaction Index used in ELC-FS is the 4-item version, and is asked in each parent's interview (PI, AI, OHP). The first item (**relsat1**) is scored on a 7-point scale, and the remaining 3 items are scored on a 6-point scale. The total index is scored by summing all scores together, with possible scores 0-6 for relsat1 and 0-5 for the other items (**dvcsi4**).

Variable name	Variable label
relsat1	How happy is respondent's relationship with their partner
coup2	Warm and comfortable relationship with partner
coup3	How rewarding is relationship with partner
coup4	Satisfaction with relationship
dvcsi4	DV Couple Satisfaction Index (CSI-4)

4.3.10 CASI module: Parental Stress

Parental Stress Scale (Berry & Jones, 1995)

The measure of parental stress used in ELC-FS was the Parental Stressors Subscale of the Parental Stress Scale (Berry and Jones, 1995). Six items on this subscale are rated on a 5-point scale, from 'strongly disagree' (1) to 'strongly agree' (5), meaning a sum of scores from 6–30 (**dvpss)**, with higher scores relating to higher levels of stress.

Variable name	Variable label
bon7	Parental Stress: child takes time and energy
bon8	Parental Stress: worry whether doing enough

Variable name	Variable label
bon9	Parental Stress: child is major source of stress
bon10	Parental Stress: Little time and flexibility
bon11	Parental Stress: child financial burden
bon12	Parental Stress: difficult to balance responsibilities
dvpss	DV Parental Stress Scale

4.3.11 Childcare module: Service use

Client Socio-Demographic and Services Receipt Inventory – European Version, CSSRI-EU (Chisholm et al., 2000)

The CSSRI-EU was developed to be a cross-cultural instrument to collect data on service utilisation, particularly those with mental disorders, for the basis of calculating costs of care (Chisholm et al., 2000).

The ELC-FS used the service receipt sub-section CSSRI-EU as the basis for collecting information on services and professionals consulted by families. The adapted version of the services receipt inventory was developed by the Children of 2020s after consultation with their funder (Department for Education) and other policy stakeholders about important services for families of young children. They developed a split list of services which ELC-FS used: those relating to specific professionals and those that are support services. For each list, respondents were asked whether they had used the service (yes/no) and if so, how many times since their child's birth.

These questions were only asked to PIs.

Variable name	Variable label
seuse1_a	Which professionals seen since birth
seuse1fr	How many times seen professional since birth
svcuse_a	DV Support services used since birth
seuse3fr	How many times used service since child's birth

5. Survey Research Data

5.1 Licensing and data access

The ELC-FS survey research data have been processed by CLS and supplied to the UK Data Service (UKDS).

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

Please refer to section 5.14 for information on how these data have been de-identified for sharing.

Safeguarded data (EUL)

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

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

Controlled data (Secure Access, SA)

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

1. Parents' gender[1]
2. Year of emigration to the UK and country of birth

[1] Only sex at birth (variables with 'birthsex' in the name) is available for 1) respondents, 2) co-resident parents of the cohort member as reported by the PI, 3) co-resident partners of the PI/OHP, 4) the OHP as reported by the PI in the EUL data to minimise the sensitive nature of data comparisons between gender and sex at birth. Gender for the different respondents/parents/partners (variables with 'gender' in the name) is available in the secure access version of the dataset.

3. Languages spoken at home
4. Full ethnicity classifications for the respondent, cohort child and co-resident partners
5. Full list of religious affiliations
6. Child's due date and whether child born early/late
7. Fertility treatments
8. Child's longstanding and developmental health conditions
9. Data from triplet families
10. Full SOC2020 and SIC codes
11. Pregnancy history of the biological mother
12. When child came home from hospital
13. Some specific dates about when parental leave started/ended
14. Number of rooms in the household
15. Respondent's age/year of birth and other parent's age/year of birth
16. Lengths of cohabitation and relationships between parents/partners
17. Whether child is in contact with their biological family if they do not live with them
18. Number of grandparents
19. Detailed information on parent-child relationships
20. Details about more rare forms of childcare in the sample (e.g. Au pairs, specific family members) and age of the child when they used different childcare
21. Whether respondent used drug and alcohol support services
22. The language the interview/translated materials used in the interview
23. Total number of people in the household including cohort members

A de-identified version of some of these variables has been created to be able to share the information as safeguarded data (EUL), e.g. outliers are grouped together - see section 5.14 for more information.

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

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

5.2 Datasets and data structure

The ELC-FS survey research data consists of two long format (hierarchical) datasets in this first deposit. Long or hierarchical (or stacked) datasets are datasets that contain multiple rows per group. In this study the group is the birth event / family that groups individuals together. For example, there are multiple rows per family because there are more than one parent/carer interview per family, or more than one child/cohort member per family.

The survey data are presented in two separate datasets, which have been structured differently depending on whether the data is displayed either by parent/carer or by the child/cohort member. The child/children is/are referred to as the cohort member(s)/cohort child(ren) as the cohort study will follow these children over time. Each cohort child is associated with at least one parent that took part in ELC-FS, and conversely each parent is associated with their cohort child (or children the in case of twins or triplets).

Both datasets contain the same content, but they are restructured either to display one row per parent/carer or one row per cohort member:

Parent dataset: One row per parent/carer (long format), thus displaying 1 to 3 parents per birth event (or family)[2], with the information about the cohort member(s) presented in a wide format, that is, on each parent row there are separate sets of variables for each cohort member (variable suffices: _cm1, _cm2, _cm3).

Cohort member dataset: One row per cohort member (long format), thus displaying 1 to 2 cohort members per birth event (or family) in the safeguarded data (EUL), and 1 to 3 cohort members in the controlled datasets (Secure Access) that also hold data from the triplet family. The information about the parents/carers is presented in wide format, that is, on each child row there are separate sets of variables for each parent/carer (variable suffices: _PA, _PB, _PC). More information on these parent labels can be found in section 5.6.

This means that there is no obvious need to merge the datasets to each other, since they contain the same information from the survey, but in different formats. The user can therefore choose which dataset to use depending on whether they want to answer a research question about the parents/carers or a research question about the children.

Because of the data structure, some information relating to the parent, their household or their children collectively is repeated per rows in the dataset structured by cohort member (one row per child * _bycm). For instance, where there are twins, the parent response for a question about the household like housing tenure (variable 'tenure') will be the same for each twin cohort member, and therefore applies to each cohort member within the family. On the converse of this, the information relating specifically to a cohort member will not be repeated across the cohort member rows because it is specific to only one cohort child of the multiple e.g. variable 'cry1'. These kinds of child specific variables have 'cm' in the variable name to help identify which child they relate to in a multiples family.

The same applies for the dataset structured by parents/carers (one row per parent/carer *_byparent), where variables specific to cohort members may be

[2] Note: while up to four parent interviews per child were allowed (PI, AI in main household, OHP and AI in OHP household) in the final data there were no families where all four interviews were utilised.

repeated because they are asked once per child (e.g. ‘cry1’). Again, these kinds of specific child variables have ‘cm’ in the variable name.

Table 5. List of safeguarded datasets (End User Licence)

Dataset name	Contents
elcfs_parent_main_interview_byparent	Parent dataset (1 row per parent/carer): interview about the Household, the parents/carers themselves, and the Cohort Members
elcfs_parent_main_interview_bycm	Cohort member dataset (1 row per cohort member) interview about the Household, the parents/carers themselves, and the Cohort Members

These two datasets are shared as safeguarded data (EUL) and contain the majority of the survey variables.

The Secure Access datasets have the same data structure as the safeguarded datasets of Table 5, with the respective names:

elcfs_parent_main_interview_byparent_sa,

elcfs_parent_main_interview_bycm_sa.

Some of the variables included in the Secure Access datasets are also shared under End User Licence datasets in a de-identified format (see section 5.14 Data de-identification).

5.3 Future datasets

Further data deposits have been planned to provide additional data which was not possible to deposit with the initial deposit.

The additional data will include:

- additional derived variables
- geographical Indicators dataset (controlled/secure access) including various indicators obtained via postcode linkage with the ONS Postcode Directory

- household dataset, a long file that contains information about relationship of the Household members to the cohort member(s) and the respondents, and the relevant derived variables.

5.4 Data documentation

In addition to this User Guide, the following documentation accompanies the data shared via the UKDS:

Table 6. Data documents

Name of the document	Content summary
Early Life Cohort Feasibility Study Questionnaire	This document provides the questions asked in the ELC-FS (the full/long version of the questionnaire), including details on any routing, and mode specific adjustments to wording. It also notes which questions were asked in the online follow-up.

5.5 Identifiers

5.5.1 Identifiers used for birth events, households and individuals

The following identifiers have been added to the datasets to help the users distinguish between birth events (the family), households (within birth events), and individuals (parents and cohort members):

Table 7: Identifiers included in ELC-FS data deposit

Variable name	Variable label	Details
elcbirthid	ELC-FS Birth Event ID shared by parents/carers and cohort member(s)	The elcbirthid starts with 'E' and is a combination of numbers and letters, for example, E10001A. A birth event is defined by a singleton birth or multiple birth 'event', e.g., twin cohort members and their parents/carers would share the same birth event ID. The elcbirthid therefore identifies all individuals associated with a birth event across households.
elcpid / elcpid_cm	ELC-FS Unique person ID (for parent or cohort member)	Parents/carers' person ID (elcpid) is a concatenation of the birth event ID with a number unique to that person, for example: A10001A_001 A10001A_002 A10001A_003

Variable name	Variable label	Details
		The number suffix aligns with the parent order ID given (PA is 1, PB is 2 and PC is 3). Cohort member's person ID (elcpid_cm) is a concatenation of the birth event ID with a number that signifies the order of the cohort member within the group of cohort member(s), for example: A10001A_cm11 for a singleton cohort member, where 'cm11' indicates one cohort member out of one in total A10002A_cm12 for a twin cohort member, where 'cm12' indicates this is the first cohort member out of two in total A10002A_cm22 for a twin cohort member, where 'cm22' indicates this is the second cohort member out of two in total The first digit of these IDs match the first digit of the suffix on the child variables (cm1, cm2, cm3)
elchhid	ELC-FS Household number within an ELC birth event – please note this is unvalidated and may be updated when household grid data is deposited	The parents/carers are distributed within one or two households (the PI or the OHP household if applicable). This household number shows whether the parent/carer is in the first or the second household. Household 1 is always the PI household; Household 2 is always the OHP household. Some birth events (15 in total) only have a household 2, as there were no interviews in the PI household for this birth event. Please note that because the household grid data has not yet been validated and deposited, the household ID numbers may change after they are checked against the household grid information for a small number of cases.

5.5.2 Additional variables to help sample selection

The parent dataset contains 1-3 parents per birth event, and the cohort member dataset contains either 1-2 cohort members per birth event in the safeguarded (EUL) data or 1-3 cohort members per birth event in the Secure Access data.

Additional variables have been added to the beginning of both datasets to help the data user identify which sub-sample they would prefer to work on:

- **numbaby** indicates the multiple birth status by showing the number of babies in the birth event, e.g. singleton=1, twins=2
- **parent_order** (PA, PB, PC) shows the order of the parent/carer respondent as listed in the parent data set, and therefore also notes each parent respondent within the birth event. The order has been determined by providing priority to parent respondents who have provided the most information (usually the PI>OHP>AI>OHPAI). More information on this is in section 5.6.

5.6 Parent dataset – one row per parent/carer

The parent dataset contains the information provided by the participants in a format where each parent/carer's response occupies a row.

This includes questions about themselves (e.g., parental health, employment) and in some cases questions about the Household and about the Cohort Member(s) of the birth event.

Some questions ask about each cohort child in a multiple birth separately, and these responses are in wide format, namely, one variable per cohort child. These variables have the suffix _cm1 _cm2 _cm3 (_cm3 is only available in Secure Access). Where there is a singleton birth, only those with suffix _cm1 will have values. These suffixes match the first digit of the elcpid_cm suffix (e.g. _cm12 is the first cohort child (_cm1) of a two cohort child family) for the cohort members.

The visualisation below shows some key characteristics of the dataset:

- The Birth Event ID (**elcbirthid**) is shared by multiple parents/carers.

- The Person ID (**elcpid**) is unique for each parent/carer, and it points to the specific respondent.
- The parent flag (**parent_order**) notes each parent respondent within the birth event. The parent flag matches the suffix of the parent data in the cohort member dataset, and aligns with the number suffix of the elcpid (i.e. PA is 1, PB is 2 and PC is 3, also denoted by the suffix of the elcpid).
- The hhid is shared by parents/carers who live in the same household, and differs for parents who live in separate households.
- The responses about each cohort member are provided in a wide format with the suffixes _cm1 and _cm2 in EUL, and _cm1, _cm2 and _cm3 in Secure Access, for the respective cohort member (these suffixes match the first digit of the ELCPID suffix (_cm12) for the cohort members).

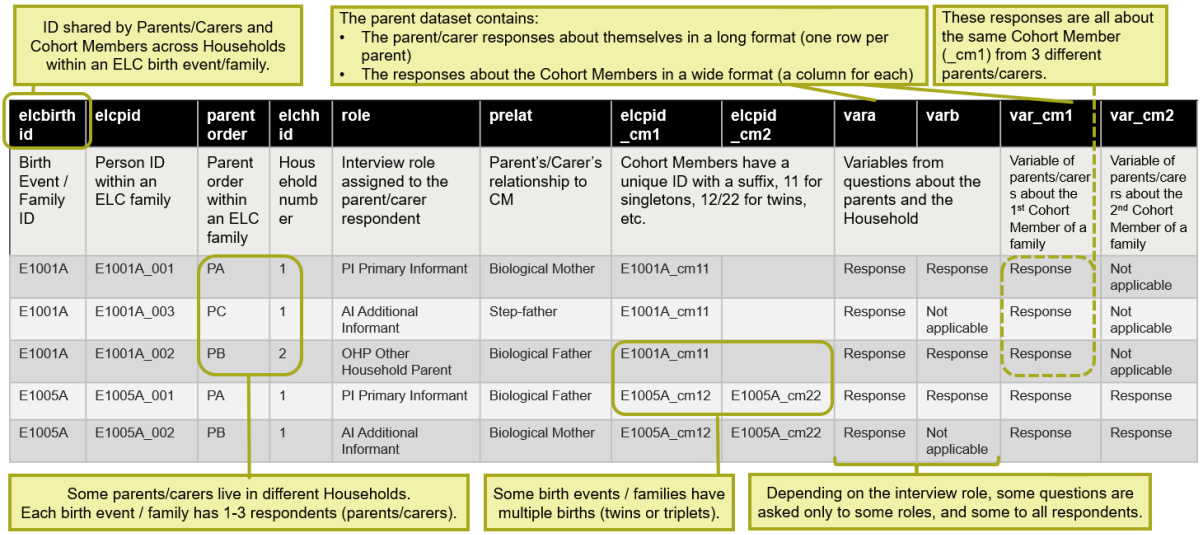


Figure 2: Visualisation of the parent dataset (one row per parent/carer)

The parent order flag options are PA, PB and PC. PA stands for 'Parent A', PB for 'Parent B' and PC for 'Parent C'. These are the parents sorted into an order, and these order labels are carried forward into the cohort member dataset so each parent's data can be tracked across the two datasets. Each child will always have a parent A complete (as every child has at least one responding parent), and parent B will be complete if there is a 2nd parent interview, and parent C will be complete if there was a 3rd parent interviewed. This ordering was done approximately by

amount of information given (with those with the more information being ordered above those with less information). The ordering within a birth event was:

1. Primary Informant interview: in cases where there two primary informant interviews for a birth event then the biological mother was prioritised as Parent A.
2. Own household parent interview
3. Additional Informant in the child's main household interview
4. Additional Informant in the OHP household interview

The distribution of the parent roles across the parent orders is as follows in the parent dataset:

Table 8: Distribution of parent roles (interview types 'parent_role') across Parent A/B/C classification ('parent_order')

	PI	OHP	AI	AIOHP	Total
Parent A	1902	16	14	0	1932
Parent B	7	44	1137	3	1191
Parent C	0	0	0	1	1
Total	1909	60	1151	4	3124

This is based on the End User Licence data by crosstabulating the variables parent_order and parent_role on the parent/carer level dataset. Total number of parent interviews (N=3124) corresponds to the number of rows in this dataset.

5.6.1 Perusing the figures of unique parents and cohort members in the parent structure dataset

How to calculate the number of unique parents/carers?

This is the total number of rows of the dataset, because the structure of the dataset is one row per parent/carer. For example, if you run frequencies of the variable 'parent_order', you will find:

In elcfs_parent_interview_byparent (EUL): 3124 parents/carers.

In elcfs_parent_interview_byparent_sa (SA): 3126 parents/carers.

How to calculate the number of unique cohort members

In the *_byparent dataset, the cohort members are distributed in a wide format, however, at the same time are repeated in multiple rows due to the fact multiple parents/carers have provided data (see 5.2 Datasets and data structure about the repeated information).

One can calculate the total number of unique Cohort Members using the *_byparent dataset by following the steps:

Step 1) selecting one row per birth event / family, for example, by selecting the rows where parent_order equals ‘PA’ since all birth events / families have at least one parent (parent A), and,

Step 2) run frequencies on a set of cohort member variables, for example, csex1, csex2 (and csex3 if using the SA dataset). Note that this will be the cohort member’s sex provided by the parent in position A (where parent_order equals ‘PA’).

Step 3) sum the valid information (non-missing), for example:

In elcfs_parent_interview_byparent (EUL): in total, there are 1973 unique cohort members, because there are: 1932 cohort members at position 1 (singleton or first of twins/triplets), 41 cohort members at position 2.

In elcfs_parent_interview_byparent_sa (SA): in total, there are 1976 unique cohort members, because there are: 1933 cohort members at position 1 (singleton or first of twins/triplets), 42 cohort members at position 2, and 1 cohort member at position 3.

5.7 Cohort member dataset - one row per cohort member

The cohort member dataset contains the information in a long format where each child is in a row. Each row contains all the responses of the parents/carers in a wide format (different sets of variables for each parent). This includes the responses about parents/carers themselves, about the household, and about the cohort member of that row.

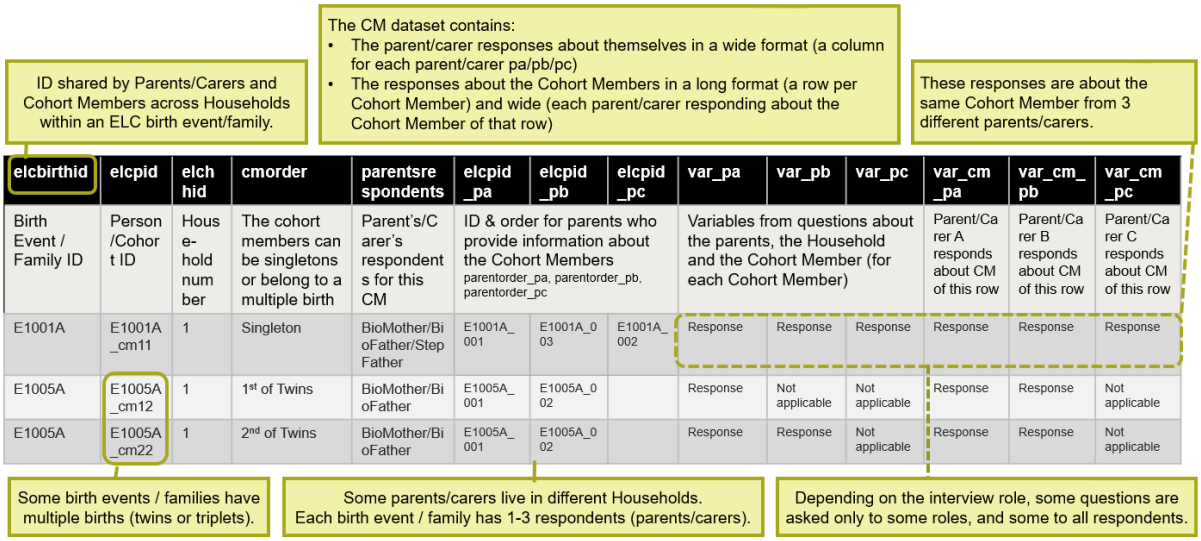


Figure 3: Visualisation of the cohort member dataset (one row per cohort member)

The visualisation above shows some key characteristics of the dataset:

- The Birth Event ID (**elcbirthid**) is shared by multiple cohort members.
- The Person ID (**elcpid_cm**) is unique for each cohort member, and the last 2 digits show whether the cohort member is singleton (11) or member of multiple birth (e.g., 12 first of twins).
- There are two variables to assist users to identify children within multiples 1) **elcfs_multiples_a** which identifies child 1 or 2 (or 3 in the Secure Access dataset) and **elcfs_multiples_b** which identifies whether the child is a singleton, first of two twins, second of two twins (or also first of triplets, second of triplets or third of triplets in the Secure Access dataset).
- The parent interviews per birth event variable (**dv_birth_event_parent_roles**) shows what parents/carers have responded for this cohort member, and the parent interviews per household variable (**dv_hh_parent_roles**) shows what parents/carers have responded for this cohort member within the same household.
- The responses of the parents/carers are in wide format with the suffixes _pa, _pb, _pc; the responses from each parent are provided together in blocks, first all the responses of _pa, then of _pb and finally of _pc. These suffixes match the **'parent_order'** flag in the parent-level dataset.

- In cases where there are only 2 parent respondents, the variables with _pc will not contain data; in cases where there is only 1 parent respondent, variables with suffix _pb and _pc will not contain data.

In this dataset, responses from the parents (about themselves, the household or the children collectively) are repeated on each of the separate rows for each cohort member of a multiple birth (e.g. twins/triplets). Responses from parents to questions which were repeated for each cohort member of a multiple birth are different across the separate rows of the data.

5.7.1 Perusing the figures of unique parents and cohort members in the cohort member structure dataset

How to calculate the number of unique cohort members?

This is the total number of rows of the dataset, because the structure of the dataset is one row per cohort member. For example, if you run frequencies of the variable 'numbaby' or 'samplecountry', you will find:

In elcfs_parent_interview_bycm (EUL): 1973 cohort members.

In elcfs_parent_interview_bycm_sa (SA): 1976 cohort members.

How to calculate the number of unique parents/carers?

In the *_bycm, dataset, the parents/carers are distributed in a wide format (in different variables '_pa', '_pb', '_pc') for each row the represents a unique cohort member. Since the *_bycm dataset contains the cohort members from multiple births (twins in SA and EUL, and triplets in SA) it means the information that the parents/carers provided may be repeated per child, for example, variables about tenure or employment (see 5.2 Datasets and data structure about the repeated information).

One can calculate the total number of unique parents/carers using the *_bycm dataset by following the steps:

Step 1) selecting one row (child) per birth event / family, for example, by selecting where the variable 'elcfs_multiples_*' equals 1, this selects the singleton or the first cohort member of twins/triplets, and,

Step 2) run frequencies on a set of parent/carer variables, for example, the variables 'parent_order_pa', 'parent_order_pb', and 'parent_order_pc'.

Step 3) sum the valid information (non-missing), for example:

In elcfs_parent_interview_bycm (EUL): in total, there are 3124 unique parents/carers, because there are: 1932 parents at position A, 1191 parents at position A, 1 parent at position C.

In elcfs_parent_interview_bycm_sa (SA): in total, there are 3126 unique parents/carers, because there are: 1933 parents at position A, 1192 parents at position A, 1 parent at position C.

5.8 Similarities and differences with the Millennium Cohort Study data

The ELC-FS has similarities to the Millennium Cohort Study (MCS), as in both studies there are multiple parents/carers providing information about the cohort member(s), and multiple cohort members per birth event (such as twins and triplets).

The ELC-FS identifier **elcbirthid** points at a group of persons that are associated with that birth event, for example, parents/carers and cohort members, just like the **mcsid** is the MCS family identifier.

Some key differences between MCS and ELC-FS are described in the table below.

Table 9: Differences between ELC-FS and MCS sweep 1 datasets

Issue	ELC-FS	MCS sweep 1
Number of parents interviewed	Up to 4 parents/carers can be interviewed for each birth event, though in the final data the maximum number achieved per birth event was 3.	Up to 2 parents/carers can be interviewed per family
Number of households per family	Each birth event might contain interviews with parents/carers from 1 or 2 households in cases where the biological parents live separately	Each family contains interviews from parents/carers from 1 household
Contents of the parent dataset	Responses of the parents/carers about themselves (1 row per parent) and	Responses of the parents/carers about themselves. The

Issue	ELC-FS	MCS sweep 1
	about the cohort members (in a wide format for the cohort members).	information about the cohort members is in separate datasets (parent_cm structure, and cm structure).
Contents of the cohort member dataset	Responses of the parents/carers about themselves (1 row per cohort member) and about the cohort members (in a wide format for the parents/carers).	Responses provided by only one parent about each cohort member. In later sweeps when the Cohort Member (cm) became a data provider, the cm structure includes the responses of the cohort member about themselves,

5.9 Variable description

Variable order

The order in which variables appear in the datasets broadly follows the order of modules, and of questions within modules, in the survey questionnaire. The order only differs when groups or sets of questions about an individual parent are blocked together in the cohort member dataset: the responses from each particular parent are blocked together (Parent A, Parent B, Parent C) in questionnaire order for each parent in the cohort member dataset. The variables specific to a cohort member (variables with cm in the name) are integrated into these blocks in questionnaire order. The parent dataset runs in questionnaire order followed by derived variables produced after fieldwork (variables beginning with dv) and weights.

Variable names

The variable names are usually the same as in the questionnaire documentation, except for the DVs created by CLS after fieldwork which are instead documented in the appendix of this user guide. This may not exactly match in all cases because of some of the reasons listed below. In the questionnaire documentation accompanying

this user guide you will notice many variables have '@' at the end of their name, which is not present in the data variable names. This '@' signifies this variable was part of the online follow-up survey.

Some common markers across variable names in the data are:

- All derived variables produced by CLS post-fieldwork have 'DV' at the start of variable name (see appendix).
- For multi-coded variables, where a single question produces more than one response, a suffix has been used to identify the iteration. 01, 02, 03….. have been used to denote the 1st, 2nd, 3rd,...iterations respectively. Please note 001 in a variable name, however, signifies not applicable (see section 5.11 on missing values).
- Where respondents were allowed to write in their own answer (when option 'OTHER' was given) variables containing this have an '_o_' in their variable name (standing for 'other').
- In most cases, these written responses were evaluated and then recombined with the original response options if they were closely aligned, or new categories were created to accommodate answers appearing in the text. These variables are marked with an '_a_' in their name to signify this has happened.
- Variables also had to be adjusted by the CLS data team to make them suitable for the safeguarded data deposit (EUL). Where variables were adjusted in the EUL version (e.g. response options for a variable were combined because of small numbers) the variable name will contain '_rec_' standing for 'recoded'. If the variable name contains '_tr' this means the variable has been truncated.

Variable labels

The variable labels included in the dataset are based on the question wording that can be found in the questionnaire documentation. Where necessary, labels have been modified in an effort to ensure they are comprehensible and accurate. The questionnaire documentation should be used for the wording of the question that collected the relevant variable.

Any recoded variables (as described in the section above) in the EUL deposit will have a description of the edits made in their variable label with the instruction ‘check SA’ to signify the user should enquire about the Special Access version of the variable to see how it was originally coded.
Related to variables mentioned above with ‘_o_’ or ‘_a_’ in their variable name, their variable label will also have an ‘(O)’ or ‘(A)’.

All derived variables, whether computed in the script or post-fieldwork by CLS, have DV at the start of the variable label.

Where one option of a multi-coded variable was an exclusive code (i.e. if it was selected then no other multi-code options can be selected) this is illustrated in the variable label as ‘[EXCLUSIVE CODE]’.

Value labels

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

5.10 Derived variables

Several derived variables have been produced by CLS based on the questionnaire data following fieldwork. Detailed documentation on their derivation can be found in Appendix 1.

Derived variables in the dataset that were produced by CLS are given the prefix ‘DV’.

5.11 Missing values

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

-9 = Refusal

-8 = Don't Know (or ‘insufficient information’ for derived variables only)

-1 = Item not applicable

-96 = Missing values (derived variables only)

The value -1 is used for missing responses to questions which participants would not have been asked if they only partially completed the survey or they were not routed to this question. Moreover, due to the data format, some variables have -1 due to data not existing, for example, columns with the suffix _cm2 and _cm3 where a second and third cohort member does not exist (family/birth event with a singleton), or columns with the suffix _pb and _pc where a second and third parent/carer does not exist (only one parent/carer respondent, not necessarily single parent family).

Where multidcoded variables have been divided into individual items, items that include all those with a particular missing value have that missing value in their name proceeded by two zeros (i.e. _009 for a refusal, _008 for don't know and _001 for not applicable).

Value -96 is only used for CLS derived variables. '-96' denotes where the score/value is missing entirely for the derivation. -8 is used for CLS derived variables to signify where there is insufficient information to derive a score (e.g. if all items of a scale need a valid answer to derive a summary score but some items have missing values). A description of the CLS derived variables is in Appendix 1.

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

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

5.13 Missing data due to routing errors

During the data editing and cleaning process, routing errors, resulting in missing data was discovered in three variables, noted below:

- The variable '**whyt**' was initially coded in the script as single coded, but half-way through fieldwork the variable was updated to be multi-coded.
- Following an error in the specification of the online follow-up survey questionnaire, '**relsat1**' (how happy is respondent's relationship with their partner) was not included in the first administrations of the online follow-up survey, and added later. This has resulted in missing relsat1 information for 85 of 130 parents in the online follow-up survey.
- In the Additional Informant (AI) questionnaire, there was an error in the routing for questions related to pregnancies due to an error in deriving '**biomum**'. This was spotted early in fieldwork and corrected. We estimate data loss of pregnancy history information for about 50 respondents.

5.14 Data de-identification

In addition to the pseudonymisation (use of anonymised IDs), the data have been examined for sensitive topics and disclosive information, as well as for rare responses (low counts), and the data have been distributed to safeguarded (EUL) and controlled (SA) datasets.

Sometimes information is too disclosive or sensitive to be shared in the EUL version. In these cases, the complete response in those variables is made available under Secure Access (SA), and sometimes a de-identified version is created and released as safeguarded data (EUL version).

The following methods have been used to de-identify variables that can be shared under EUL. Certain suffixes have been used in variable names to highlight that they have been de-identified:

1) **Truncation**: the truncated variables in the EUL version are named with suffix **_tr.** This has been applied to Socio Economic Codes (SOC/SIC).

2) **Recoding**: the recoded variables in the EUL version are named with suffix **_rec**. This has been applied in different manners depending on the variable contents:

- Grouped in the upper and/or lower end of the distribution where the values have low counts (e.g., length of relationship, number of previous pregnancies).
- Recoded certain values that contain sensitive or disclosive information.
- Creation of a flag: a variable with the suffix **_flag** indicates whether any of a group of variables contains a response. In the meantime, the flag provides the information on whether at least one of the conditions in that group is mentioned (e.g., any type of a health classification) and it is available on the EUL version. All the variables from that group are available as controlled data under Secure Access.
- For certain potentially disclosive multi-coded data, such as which relatives look after the child (ccrel_a_06 to ccrel_a_10), low-count responses have been combined into a new variable labelled 'Other' (ccrel_a_05_rec). This variable combines all coded and back-coded categories not shared as safeguarded data under EUL, while the full breakdown is available as controlled data under Secure Access.

3) **Removal**: all text variables that contain verbatim information provided by the respondents have been removed from both the EUL and Secure Access research datasets. This includes job titles, job descriptions, exact names of education institutions, town name, postcodes and the final open-ended question. These potentially identifiable CLS data can be accessed securely by applying directly to the CLS Data Access Committee.

Details about variables that have been put into secure access, and whether a de-identified version of the variable been included in the end user license version, can be found in the table below.

Table 10: Variables that have been removed from EUL and put into secure access, and whether a deidentified version of the variable is available in EUL.

Type of variable not available in EUL	Variables in Secure Access only	De-identified version of variable available in EUL
Parents' gender	pgender; ohpgender	

Type of variable not available in EUL	Variables in Secure Access only	De-identified version of variable available in EUL
Year of emigration to the UK and country of birth	bircou; migy	bircou_rec; migy_rec
Languages spoken at home	langho	langho_flag
Respondent's, cohort child's and partner's ethnicity	pethnic_a, ethnic2_a, ethnic_a	dvcmethnic6, dvcmethnic8, dvcmethnic11, dvethnic4, dvethnic6, dvethnic8, dvethnic11, dvethnicp4, dvethnicp6, dvethnicp8, dvethnicp11
Religion	relig_a	relig_a_rec
Child's due date and whether child born early/late	pregb1m; pregb1y; pregb2; pregb3; pregb4; pregb4a	
Fertility treatments	desem; trwt; ftrl	
Child's longstanding and developmental health conditions	chlhp_cm	chlhp_cm_flag
Data from triplet families	all variables with c3 or cm3 in the name	
Full SOC2020 and SIC codes	soc2020, sic	soc2020_tr, sic_tr

Type of variable not available in EUL	Variables in Secure Access only	De-identified version of variable available in EUL
Pregnancy history of biological mother	cmpregchk; pregmany; preglbn	pregmany_rec; preglbn_rec
Date child came home from hospital	pregb6	pregb6_rec
Some exact days parental leave started and ended	workstop4d; leave2pad	
Number of rooms in the household	numrooms	numrooms_rec
Respondent's age/year of birth and other parent's age/year of birth	p_age; ohpage; pdoby; ohpdoby	p_age_rec; ohpage_rec; pdoby_rec; ohpdoby_rec
Lengths of cohabitation and relationships between parents/partners	hbioplivey; hbioplivem; hnbppcoupley; hnbppartlivy; ohprelly;	hbioplivey_rec; ohprelly_rec
Whether child is in contact with their biological family if they do not live with them	birthp; birthpc; nrbiop	nrbiop_rec
Number of grandparents	gali	gali_rec
Detailed information on parent-child relationships	dv_prelat_tocm; biomum; biodad; dvhhcarers	dv_prelat_tocm_rec; dvhhcarers_rec
Details about more rare forms of childcare and	ccar_a_05; ccrel_a_06; ccrel_a_07; ccrel_a_08; ccrel_a_09; ccrel_a_10;	ccar_a_04_rec; ccrel_a_05_rec; csag_rec; cage_rec;

Type of variable not available in EUL	Variables in Secure Access only	De-identified version of variable available in EUL
age child was when they used different childcare	cage; cend_05; csag; chou_05; cday_05	cend_04_rec; chou_04_rec ; cday_04_rec
Whether respondent used drug and alcohol support services	svcuse_a_15; seuse3fr_15	
Language of interview/translated materials used in interview	oeco; oetm	
Total number of people in household	dvnumall	dvnumallhh_rec

5.15 Output Disclosure Control (for controlled data)

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

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

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

6. Design and non-response weights

6.1 Design Weights

Design weights allow data users to account for the sample design of a survey by reweighting the sample to the population from which it was drawn. As noted in the section on sample design (2.1.2), Scotland, Wales and Northern Ireland were oversampled relative to England. Analyses of data across multiple UK countries therefore need to be weighted to account for this. The sample design also included two additional boosts in England only: an ethnic minority boost of Black African and Black Caribbean babies and Pakistani and Bangladeshi babies, and an area-based low-income boost. Analyses of data from England therefore also need to be weighted to account for this.

Separate country-specific analyses within Wales, Scotland or Northern Ireland do not require a design weight, so all respondents in these countries have been given a country-specific design weight of 1. This ensures that design weighted country-specific analyses of all respondents within each of these countries have the desirable property that the effective sample size is equal to the achieved sample size (the total number of responders). For analyses of data from England, the design weights are calculated as the inverse of the selection probability (i.e., the total population size divided by the number selected for the study), resulting in group-specific design weights of 11.9 for children of Black African or Caribbean ethnicity, 14.7 for children of Pakistani or Bangladeshi ethnicity, 22.5 for children living in a deprived PSU and not in one of the above ethnic groups and 66.2 for children not living in a deprived PSU and not in one of the above ethnic groups. These design weights are then scaled so that the weighted sample size of weighted analyses of all respondents in England is equal to the number of respondents in England. These country-specific design weights are supplied with the dataset as the variable **w_design_cs_scaled.** These weights are provided for country-specific analyses.

To account for the oversampling of Scotland, Wales and Northern Ireland relative to England, a further set of weights are provided for UK-wide analyses. These have been scaled so that the weighted sample size in each country is in proportion to the

total population of eligible births in that country but the total weighted sample size is equal to the achieved sample size across the UK. These UK-wide design weights are supplied with the dataset as the variable **w_design_uk2.**

6.2 Non-response Weights

6.2.1 Introduction

Some degree of non-response is inevitable in all surveys. Non-response means less statistical power but can also introduce bias as respondents often differ systematically from non-respondents. Here we briefly describe the derivation and implementation of non-response weights that are provided with ELC-FS data. While the non-response weight is not provided in this deposit, it is used to create a combined weight with the design weights above.

6.2.2 Response Definition

For the purpose of non-response weight derivation, response is defined as any interview response in the child's primary household (i.e. Primary Informant (PI) or Additional Informant (AI), including partially completed as well as fully completed interviews. This aligns for our definition of productive families for the purposes of response rate calculation (discussed in section 3). This means that the small number of families in which the only interview response was outside the child's primary household (i.e. from an Own Household Parent (OHP)) are considered as non-respondents for this purpose.

Non-response weights are derived for respondents (by the above definition) in the achieved sample after the implementation of the online follow-up survey (i.e., responses across the main fieldwork and online follow-up surveys combined), including data subsequently obtained from re-interviews in England and Wales due to the fraudulent interviewer. Birth events in which the only interview response was from an OHP household (i.e. OHP only, or OHP and AIOHP only), which as noted above are considered as non-respondents for this purpose, will therefore not have a derived non-response weight nor a combined weight.

6.2.3 Derivation of Non-Response Weights

One significant advantage of the ELC-FS sampling frames is the amount of information available for both respondents and non-respondents, which can be utilised in the derivation of non-response weights.

We derived non-response weights for England, Wales, Scotland and Northern Ireland separately due to differences in sampling frame variable availability and data access restrictions. Access to the whole population, or alternatively the entire selected sample (i.e., prior to National Data Opt-out and data holder exclusions), would allow us to derive weights which would reweight the sample to the whole population. In the absence of access to these, we derived weights instead using the selected returned sample, so that respondents are reweighted to this sample. As noted in Section 2.1., the 'returned sample' comprises the named sample after both stages of sampling were completed, and after data holder exclusions were applied. Additional cases were excluded later due to sensitivity or ineligibility.

In order to derive the non-response weights, binary response (as defined above) was modelled using logistic regression in terms of the sampling frame variables listed in Table 11.

Table 11. Sampling frame variables included in the non-response models.

Variable	England	Wales	Scotland	Northern Ireland
Child's sex	✓	✓	✓	✓
Child's ethnicity	✓	✓	✓	✓
Child's birthweight	✓	✓	✓	✓
Child's gestational age	✓	✓	✓	✓
Mother's age at birth	✓	✓	✓	✓
Father's age at birth	✓	✓	✓	
Mother's country of birth	✓	✓		✓
Father's country of birth	✓	✓		

Variable	England	Wales	Scotland	Northern Ireland
IMD decile	✓	✓	✓	✓
Region	✓			
Ethnicity/area deprivation group	✓			
Sampling stratum	✓			
Incentive group	✓	✓	✓	
Birth in marriage	✓	✓		
Birth informant	✓	✓		
Mother's total births	✓	✓		
Mother's total previous pregnancies			✓	
Mother's parity			✓	
Mother's occupation			✓	
Father's occupation			✓	
Mother's ethnic group			✓	✓

In England, Wales, and Scotland, the small amount of missing data on the sampling frame variables was handled using multiple imputation (MI) prior to the derivation of the non-response weights. The imputation model included all the sampling frame variables in the response model, including the response indicator. Ten imputed datasets were created using chained equations. Such a relatively small number of imputations was deemed sufficient as only point estimates (the probability of response) were to be estimated from the MI analysis (i.e., inferences were not being made). Models for response were fitted in each imputed dataset and combined using standard rules (see Tables A2.1-A2.3 (Appendix) for final response models).

The Northern Ireland dataset had no missing data in the sampling frame variables, so the response model was fitted in the observed data without application of MI.

However, the response model for Northern Ireland cannot be presented here due to restrictions caused by the low counts in some of the categories of specific variables.

Across all countries, the probability of response was predicted from the model for each respondent and the non-response weight calculated as the reciprocal of this probability. Higher weight values reflect a lower probability of response and thus a responding individual with a higher weight is used to reflect a larger section of the target population The distributions of the derived non-response weights are summarised in Table 12. The derived non-response weights were then multiplicatively combined with the design weights as necessary (i.e., when analysing England only or England and Wales in combination) to provide combined weights.

Table 12. Distributions of non-response weights.

Country	Minimum	Mean	Maximum
England	1.00	2.23	9.54
Wales	1.00	2.23	10.14
Scotland	1.00	1.86	7.06
Northern Ireland	1.00	1.98	8.56

The combined weights were then scaled so that their sum equals the achieved sample size within the country (i.e., so that a weighted analysis of all respondents in that country would have a weighted sample size equal to the achieved sample size in that country). These country-specific combined weights are supplied with the dataset as the variable **w_combined_cs**.

An additional set of weights for use in UK-wide analyses were scaled so that the weighted sample size in each country is in proportion to the total population of eligible births in that country but the total weighted sample size is equal to the achieved sample size across the UK. These UK-wide final weights are supplied with the dataset as the variable **w_combined_uk2**.

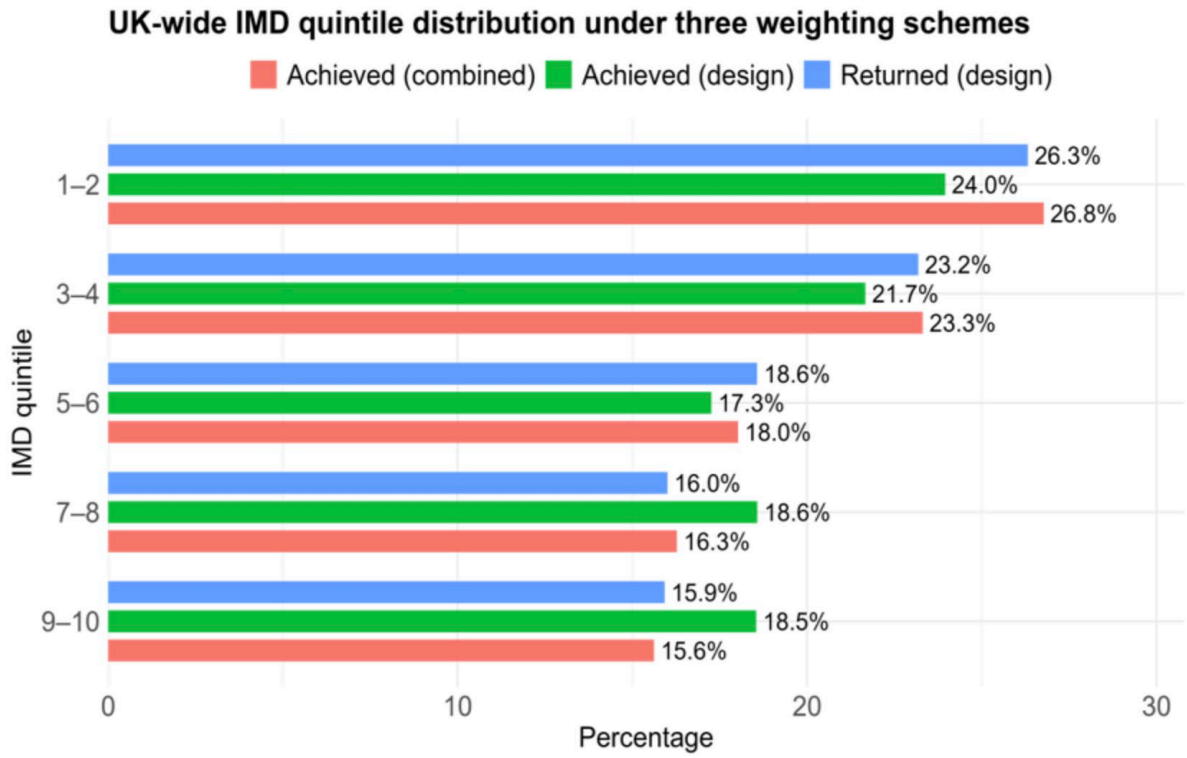
6.2.4 Performance of Non-Response Weights

The effectiveness of the non-response weights can be assessed through comparison of the combined (i.e., design and non-response) weighted distributions of sampling frame variables in the achieved sample with their design weighted distributions in the returned sample. The distributions of the deprivation indices and maternal age are presented in Figs. 4 and 5 for England, Wales, Scotland and Northern Ireland combined.

Due to the relatively lower response rates in more disadvantaged areas, the distribution of the index of multiple deprivation (IMD) in the achieved sample is slightly skewed towards the higher quintiles (i.e., less disadvantaged areas) compared to the returned sample (Fig. 4). For example, 18.5% vs. 15.9% in the top quintile and 18.6% vs. 16.0% in the second top quintile. The application of non-response weights helps better align the achieved sample with the returned sample, with these differences reduced to 15.6% vs. 15.9% and 16.3% vs. 16.0%, respectively.

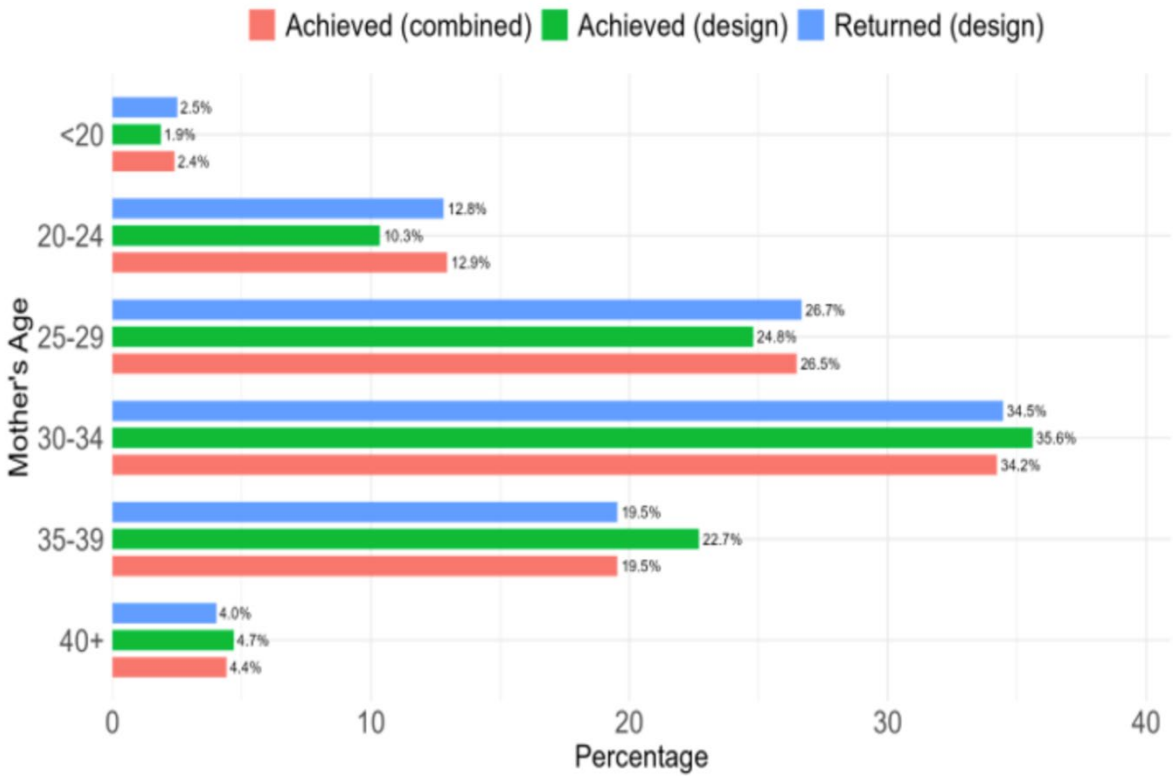
The relatively lower response rates among younger mothers similarly mean that the distribution in the achieved sample is somewhat skewed towards older mothers relative to the returned sample (e.g., 22.7% vs. 19.5% for age 35-39 for design weighted samples) (Fig. 5). The application of non-response weights removes this difference: both the achieved combined sample and returned sample are 19.5%.

Figure 4: Distribution of IMD quintiles in each sample in England, Wales, Scotland and Northern Ireland combined.



Note 1-2 = most deprived; 9-10 = least deprived. Blue = design weighted returned sample; green = design weighted achieved sample; red = combined (design and non-response) weighted achieved.

Figure 5: Distribution of mother's age at birth in each sample in England, Wales, Scotland and Northern Ireland combined.



Blue = design weighted returned sample; green = design weighted achieved sample; red = combined (design and non-response) weighted achieved.

6.2.5 Implementation of Non-Response Weights

Data users should use weights that reflect both (a) the population they want their analysis to be representative of and (b) the sample used in their analysis. The weights we have provided are designed to reweight the full achieved sample to the returned sample (as a proxy for the population of all births in absence of access to data which are truly population-representative). If the analytic sample is only a subset of the full achieved sample – for example, low birthweight babies or socioeconomically deprived families – then there is no guarantee that the shared non-response weights will perform as anticipated.

ELC-FS used a complex sampling design to recruit cohort members: at the first stage, a random sample of areas using Census geographies (primary sampling units; PSUs) was selected for each country with probability proportionate to the

number of births; at the second stage, a stratified random sample of children was sampled within each selected area. Analyses of ELC-FS data should account for this complex sampling design by specifying the PSU (variable '**w_cluster_id_combined**') and stratum (variable '**w_stratum_combined**') in addition to the final weights detailed above. In Stata, this can be achieved by first using `svyset` to specify the survey design and then conducting analyses using the `svy` prefix, e.g.:

```
svyset w_cluster_id_combined [pweight=XXXX],
strata(w_stratum_combined)
svy: proportion AAAA
```

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

```
library(survey)
elcfs_svy <- svydesign(id = ~ w_cluster_id_combined, strata = ~
w_stratum_combined, weights = ~ XXXX, data = elcfs)
svytable(~ AAAA, elcfs_svy)
```

XXXX can be filled with any of the four weight variables available depending on whether the user is doing a country-specific or UK-wide analysis. The user would need to restrict their sample to just one country for the country-specific analysis using variable **samplecountry.** AAAA can be filled with whichever variable the user is exploring.

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

For some commands, users may find that certain functionalities have not been adapted for use with complex survey data. In this case, users may consider using survey weights without declaring complex survey design and noting this in the write up of their analyses. Users can also check whether accounting for complex survey design makes meaningful difference to their particular analysis by running analyses

declaring and not declaring the complex design (in situations where both these analyses are possible).

7. Mode effects

Interviews in the ELC-FS were carried out in four different modes (Face-to-face, Telephone, Teams (video), Online). See section 3.2 for distribution of modes by informant type. A feature of mixed mode designs is the potential for responses to differ systematically between survey modes. For instance, the presentation of a survey item either orally or visually can influence responses, and sensitive information may be reported more accurately when given anonymously (e.g., by web survey compared with face-to-face interview). Differences in responses arising from differences in measurement between surveys modes (rather than due to selection of different types of respondents into modes) are termed ‘mode effects’.

Unaccounted for, mode effects can generate bias in analyses, both for descriptive and inferential statistics. For instance, estimates of the change in mental health scores may reflect differences in the survey modes used.

Simply adding an indicator variable for survey mode into analyses of ELC-FS data may not be sufficient to remove bias as selection into mode was not random; In ELC-FS Additional Informant interviews were much more likely to have been completed online because of the instructions to interviewers to more readily offer this option, and Additional Informants are also more likely to be fathers. Furthermore, participants who did not respond to initial invitations all completed a shorter web version of the survey, and likely differ on a number of dimensions from those who responded at first contact. Observed differences between modes are a combination of mode effects and selection effects. Adding an indicator variable for mode may not appropriately address this.

CLS has released user guidance on handling mode effects in its cohort studies (Wright et al., 2024). This guidance introduces frameworks for understanding the consequences of mode effects for your own research questions and describes methods for reducing bias from mode effects. It also includes worked examples in R and Stata and contains a set of recommendations that we suggest researchers follow in their own analyses of CLS data.

8. References

Ahmadzadeh, Y. I., Lester, K. J., Oliver, B. R., & McAdams, T. A. (2020). The Parent Play Questionnaire: Development of a parent questionnaire to assess parent–child play and digital media use. *Social Development*, *29*(4), 945–963.

Berry, J. O., & Jones, W. H. (1995). The Parental Stress Scale: Initial Psychometric Evidence. *Journal of Social and Personal Relationships*, *12*(3), 463-472.

Bronte-Tinkew, J., Carrano, J., Horowitz, A., & Kinukawa, A. (2008). Involvement Among Resident Fathers and Links to Infant Cognitive Outcomes. *Journal of Family Issues*, *29*(9), 1211-1244.

Brugha T, Bebbington P, Tennant C, Hurry J. The List of Threatening Experiences: a subset of 12 life event categories with considerable long-term contextual threat. Psychol Med. 1985 Feb;15(1):189-94

Chisholm D, Knapp MRJ, Knudsen HC, Amaddeo F, Gaite L, van Wijngaarden B. Client Socio-Demographic and Service Receipt Inventory – European Version: development of an instrument for international research: EPSILON Study 5. *British Journal of Psychiatry*. 2000;177(S39):s28-s33.

Funk, J.L., & Rogge, R.D. (2007). Testing the Ruler with Item Response Theory: Increasing Precision of Measurement for Relationship Satisfaction with the Couples Satisfaction Index. *Journal of Family Psychology, 21*, 572-583.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L. T., Walters, E. E., & Zaslavsky, A. M. (2002). *Kessler Psychological Distress Scale (K6, K10)* [Database record]. APA PsycTests.

Kliem S, Mößle T, Rehbein F, Hellmann DF, Zenger M, Brähler E. A brief form of the Perceived Social Support Questionnaire (F-SozU) was developed, validated, and standardized. J Clin Epidemiol. 2015 May;68(5):551-62.

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

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

Putnam SP, Helbig AL, Gartstein MA, Rothbart MK, Leerkes E. Development and assessment of short and very short forms of the infant behavior questionnaire-revised. J Pers Assess. 2014;96(4):445-58.

Sadeh, A. (2004). A brief screening questionnaire for infant sleep problems: validation and findings for an internet sample. *Pediatrics, 113*(6), e570–e577.

Seifer, R., Sameroff, A.J., Dickstein, S., & Hayden, L.C. (1996). Parental psychopathology and sleep variation in children. *Child and Adolescent Psychiatric Clinics of North America, 5,* 715-727.

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

Appendix 1: Derived Variables Guide

This Appendix describes the derived variables for the Early Life Cohort Feasibility Study by CLS. All derived variables produced by CLS have variable names beginning with ‘dv’ as well as their variable label, so they can be distinguished from any derived variables done within the script (which only have dv in their variable label). Any derived variables done within the script are not listed here and can be found in the questionnaire.

These variables are all included in the main datasets under End User Licence unless otherwise indicated in section 5.14.

The study team would like to thank the Children of 2020s study team, particularly Hannah Harding and Kelly Ward at Ipsos, for supplying syntax to derive some of these variables.

A1.1 Geography

The geographical variables are produced using postcode data collected or confirmed at interview linked with the ONS Postcode directory- February 2025 edition. One case was missing dvctry because they were missing a postcode. This one case had their country imputed in this variable using the country they were sampled from.

dvctry - “DV Country of interview”

Description: Country at point of interview.
Population: All respondents.
Value and Value Labels:

(1) “England”
(2) “Northern Ireland”
(3) “Scotland”
(4) “Wales”

dvrgn - “DV December 2020 Region of interview”

Description: Region at point of interview.

Population: All respondents.

Value and Value Labels:

(1) “North East”

(2) “North West”

(3) “Yorkshire and The Humber”

(4) “East Midlands”

(5) “West Midlands”

(6) “East of England”

(7) “London”

(8) “South East”

(9) “South West”

(10) “Scotland”

(11) “Wales”

(12) “Northern Ireland”

dvimdd - “DV IMD Overall Rank Decile”

Description: Index of Multiple Deprivation rank decile: England 2019, Scotland 2020, Wales 2019, Northern Ireland 2017.

Population: All respondents.

Value and Value Labels:

(1) “Most deprived decile”

(10) “Least deprived decile”

(-1) “Not applicable”

dvidacid - “DV English Income Deprivation Affecting Children Index 2019 Decile”

Description: Decile of proportion of all children aged 0 to 15 living in income deprived families- England.

Population: England residents.

Value and Value Labels:

(1) “Most deprived decile”

(10) “Least deprived decile”

(-1) “Not applicable”

dvru11ind - “DV 2011 Census rural-urban classification- recoded”

Description: 2011 Census rural-urban classification- recoded for disclosure control.

Population: All respondents.

Value and Value Labels:

(“EWA”) “(England/Wales) Urban major conurbation”

(“EWB”) “(England/Wales) Urban minor conurbation”

(“EWC”) “(England/Wales) Urban city and town (any setting)”

(“EWD”) “(England/Wales) Rural town and fringe (any setting)”

(“EWE”) “(England/Wales) Rural village/Rural hamlet and isolated dwellings (any setting)”

(“NIM”) “(Northern Ireland) Mixed Urban/Rural”

(“NIR”) “(Northern Ireland) Rural”

(“NIU”) “(Northern Ireland) Urban”

(“SC1”) “(Scotland) Large Urban Area”

(“SC2”) “(Scotland) Other Urban Area”

(“SC3”) “(Scotland) Small Town (Accessible/Remote/Very remote)”

(“SC4”) “(Scotland) Rural (Accessible/Remote/Very remote)”

A1.2 Paradata

dv_birth_event_parent_roles - “DV Parent interviews per birth event”

Description: Summary of the parent interview responses per birth event.

Population: All birth events

Value and Value Labels:

(1) “AI only”

(2) “AI and OHP”

(3) “OHP only”

(4) “OHP and AIOHP”

(5) “PI and OHP”

(6) “PI only”

(7) “PI and AI”

(8) “PI, OHP, and AIOHP”

(9) “PI and PI”

dv_birth_event_mainhhinterview - “Flag of whether any interviews in the child’s main household (i.e. PI or AI interview)”

Description: Summary of whether any interview achieved for birth event in child’s main household (a PI or AI interview)

Population: All birth events

Value and Value Labels:

(1) “At least one interview in child’s main household”

(2) “No interviews in child’s main household”

dv_hh_parent_roles - “DV Parent interviews per household”

Description: Summary of the parent interview responses per household.

Population: All households

Values:

(1) “AI only”

(2) “OHP only”

(3) “OHP and AIOHP”

(4) “PI only”

(5) “PI and AI”

(6) “PI and PI”

total_parents_perhh - “Total number of parent interviews per household”

Description: Total number of parent interviews within a given household

Population: All households

total_parents_perbirth - “Total number of parent interviews per birth event”

Description: Total number of parent interviews within a given birth event

Population: All birth events

A1.3 Activities and Employment

dvactivity - “DV Respondents Economic Activity Status”

Description: Respondent’s current employment status.

Population: All respondents

Value and Value Labels:

(1) "Employee - in paid work"

(2) "Self employed"

(3) "In unpaid/voluntary work"

(4) "Unemployed"

(5) "Education: School/college/university"

(6) "Apprenticeship"

(7) "On government scheme for employment training"

(8) "Sick or disabled"

(9) "Looking after home or family"

(10) "Something else"

(-1) “Not applicable”

Derivation description: The variable WORKDER is harmonised to Next Step’s Sweep 9 W9DACTIVITYC.

dvmainhhworkstatus - “DV Combined labour market status of child's main household (PI and their partner) – please note this is unvalidated and may be updated when household grid data is deposited”

Description: Combined labour market status of PI and partner

Population: All PI households

Value and Value Labels:

(1) "Both PI and partner work"

(2) "PI works, partner does not work"

(3) "PI works, partner work status unknown”

(4) "PI works, no partner"

(5) "PI does not work, partner works”

(6) "Both PI and partner do not work”

(7) "PI does not work, partner work status unknown”

(8) "PI does not work, no partner"

(9) "PI work status unknown, partner does not work"

(10) "Both PI and partner work status unknown"

Derivation description: The variables WORKDER and WORKDERP recoded into working/not working/unknown according to Next Steps Sweep 9, then combined for each PI household.

soc2020_tr - "DV Occupation code - SOC2020 (truncated for low counts)"

Description: Truncation of soc2020 variable into 2 or 3 characters for low counts.

Population: All respondents

Value and Value Labels:

(xxx) "Uncodeable"

Derivation description: Truncated using variable soc2020 supplied by Ipsos. The full code was truncated to 3 characters, unless there were low counts (under a threshold of 20 or 30) where they were further truncated to 2 characters.

sic_tr - "SIC coding (truncated for low counts)"

Description: Truncation of sic variable into 2 or 3 characters for low counts.

Population: All respondents

Derivation description: Truncated using variable sic supplied by Ipsos. The SIC codes for EUL are truncated to 2 or 3 characters. The logic is as follows: all codes are truncated up to 3 characters, and then if there are still low counts (under a threshold of 20 or 30) then those are truncated further to 2 characters. This includes the dots that were part of the SIC codes of the ELC data. The same applies to the SOC codes.

dvnssec - "DV National Statistics soc2020 Socio-economic operational classification based on 2020 guidance"

Description: National statistics soc2020 socio-economic operational classification

Population: All respondents

Value and Value Labels:

(1) "Employers in large organizations"

(2) “Higher managerial/admin. occupations"

(3) “Higher professional occupations”

(4) “Lower professional/higher technical occupations”

(5) “Lower managerial/admin. occupations"

(6) “Higher supervisory occupations”

(7) “Intermediate occupations”

(8) “Employers in small organizations”

(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) “Occupations not stated or inadequately described”

(17) “Not classifiable for other reasons”

(-1) “Not applicable”

Derivation description: This variable contains a two-digit National Statistics Socio-economic Classification (NS-SEC) code, derived from SOC2020 and some additional interview questions according to ONS guidance.

dvnssec13 - “DV NS-SEC 13 (Combined operational categories)”

Description: Respondent’s current job in 13 operational categories NS-SEC format.

Population: All respondents

Value and Value Labels:

(1) “Employers in large organizations”

(2) “Higher managerial/admin. occupations"

(3) “Higher professional occupations”

(4) “Lower professional/higher technical occupations”

(5) “Lower managerial/admin. occupations"

(6) “Higher supervisory occupations”

(7) “Intermediate occupations”

(8) “Employers in small organizations”

(9) “Own account workers”

(10) “Lower supervisory occupations”

(11) “Lower technical occupations”

(12) “Semi-routine occupations”

(13) “Routine occupations”

(-8) “Insufficient information”

Derivation description: Derived from full NS-SEC (17 categories) into 13 classes.

dvnssec8 - “DV: NS-SEC 8-Class analytic classes”

Description: Respondent’s current job in eight category NS-SEC format.

Population: All respondents

Value and Value Labels:

(1) “Higher managerial and professional occupations”

(2) “Lower managerial and professional occupations”

(3) “Intermediate occupations”

(4) “Small employers and own account workers”

(5) “Lower supervisory and technical occupations”

(6) “Semi-routine occupations”

(7) “Routine occupations”

(8) “Never worked and long term unemployed”

(9) “Not classified”

(-1) “Not applicable”

Derivation description: Derived from full NS-SEC (17 categories) into 8 classes.

dvnssec7 - “DV NS-SEC 7 analytic classes (last known job)”

Description: Respondent’s current job in seven category NS-SEC format.

Population: All respondents

Value and Value Labels:

(-9) “Prefer not to say”

(-8) “Insufficient information”

(1) “Higher managerial and professional occupations”

(2) “Lower managerial and professional occupations”

(3) “Intermediate occupations”

(4) “Small employers and own account workers”

(5) “Lower supervisory and technical occupations”

(6) “Semi-routine occupations”

(7) “Routine occupations”

Derivation description: Derived from full NS-SEC analytic sub-classes (dvnssec8) into 7 classes.

dvnssec5 - “DV: NS-SEC 5-Class analytic classes”

Description: Respondent’s current job in five category NS-SEC format.

Population: All respondents

Value and Value Labels:

(-9) “Not classified”

(-8) “Never worked and long-term unemployed”

(-1) “Not applicable”

(1) “Managerial and professional occupations”

(2) “Intermediate occupations”

(3) “Small employers and own account workers”

(4) “Lower supervisory and technical occupations”

(5) “Semi-routine and routine occupations”

Derivation description: Further simplified from dvnssec8 into 5 classes.

A1.4 Finance

dvbene - “DV Whether household in receipt of state benefits – please note this is unvalidated and may be updated when household grid data is deposited”

Description: Whether any respondent in the household receives benefits.

Population: All households

Value and Value Labels:

(1) “Receives benefits”

(2) “Does not receive benefits”

(-1) “Not applicable”

(-8) “Don't Know”

(-9) “Refusal”

Derivation description: The household is categorised as receiving benefits if they reported receiving universal credit (UNCR) or any of the benefits mentioned in BENT_01 to BENT_14.

A1.5 Education

dvnvqacad - “DV NVQ equivalent of highest Academic qualification”

Description: Respondent’s highest academic qualification categorised by the NVQ equivalent.

Population: All respondents

Value and Value Labels:

(0) “Entry level”

(1) “NVQ Level 1”

(2) “NVQ Level 2”

(3) “NVQ Level 3”

(4) “NVQ Level 4”

(5) “NVQ Level 5”

(-1) “Not applicable”

(-8) “Don't Know”

(-9) “Refusal”

Derivation description: Academic qualification (EDUC1) with the highest associated NVQ level.

dvnvqvoc - “DV NVQ equivalent highest vocational qualification”

Description: Respondent’s highest vocational qualification categorised by the NVQ equivalent.

Population: All respondents

Value and Value Labels:

(0) “Entry level”

(1) “NVQ Level 1”

(2) “NVQ Level 2”

(3) “NVQ Level 3”

(4) “NVQ Level 4”

(5) “NVQ Level 5”

(-1) “Not applicable”

(-8) “Don't Know”

(-9) “Refusal”

Derivation description: Vocational qualification (EDUC2) with the highest associated NVQ level.

dvnvqacadvoc - “DV NVQ equivalent of highest Academic or Vocational qualification”

Description: Respondent’s highest academic or vocational qualification categorised by the NVQ equivalent.

Population: All respondents

Value and Value Labels:

(0) “Entry level”

(1) “NVQ Level 1”

(2) “NVQ Level 2”

(3) “NVQ Level 3”

(4) “NVQ Level 4”

(5) “NVQ Level 5”

(-1) “Not applicable”

(-8) “Don't Know”

(-9) “Refusal”

Derivation description: The NVQ level of either the highest academic qualification (EDUC1) and vocational qualification (EDUC2) depending on which is higher.

A1.6 Health

dvbirthweightkiloscm[1-3] - “CM DV Birth weight-Kilos”

Description: CM’s birth weight in kilograms for each CM.

Population: All cohort members

Value and Value Labels:

(-99) “Insufficient information”

Derivation description: Derived from child birthweight in kilos (WEIG2KG) and in grammes (WEIG2GM), or child birthweight in pounds (WEIG3LBS) and in ounces (WEIG3OZ).

dvbirthweightouncescm[1-3] - "CM DV Birth weight-Oz"

Description: CM's birth weight in ounces for each CM.

Population: All cohort members

Value and Value Labels:

(-99) "Insufficient information"

Derivation description: Derived from child birthweight in kilos (WEIG2KG) and in grammes (WEIG2GM), or child birthweight in pounds (WEIG3LBS) and in ounces (WEIG3OZ).

dvbirthweightpoundscm[1-3] - "CM DV Birth weight-Lbs"

Description: CM's birth weight in pounds for each CM.

Population: All cohort members

Value and Value Labels:

(-99) "Insufficient information"

Derivation description: Derived from child birthweight in kilos (WEIG2KG) and in grammes (WEIG2GM), or child birthweight in pounds (WEIG3LBS) and in ounces (WEIG3OZ).

dvdisabilityea - "DV Disability classification Equality act (2010)"

Description: The classification of whether the respondent is disabled according to the Equality act 2010

Population: All respondents

Value and Value Labels:

(0) "Not disabled (Equality act)"

(1) "Disabled (Equality act)"

(-8) "Don't know"

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

dvduedatedifference - "DV Due date difference in weeks"

Description: Difference between the due date and birth date of CM, in weeks.

Population: All cohort members

Value and Value Labels:

(-888) "Don't know"

Derivation description: Calculated using either number of days early (PREGB4) and whether baby early/late/on time (PREGB2) as reported by PI, or derived using due date (PREGBD, PREGBM, PREGBY) and birth date (CDOB or CDOBD/CDOBM/CDOBY). Negative values mean CM was delivered earlier than due date. The number of days is then converted into weeks.

dvduedatedifference_flag - "DV Due date difference – flag for implausible difference"

Description: This flags cases where DVDueDateDifference are implausible, suggesting incorrection information by respondents, or data entry errors by interviewers.

Population: All cohort members

Value and Value Labels:

(0) "Plausible"

(1) "Implausible"

Derivation description: If DVDueDateDifference is less than -18 or greater than +4, then it is considered implausible.

dv_ibq_cm1, dv_ibq_cm2, dv_ibq_cm3 - "DV CM Average score of Infant Behaviour Questionnaire scale"

Description: The average score of IBQ1-14 where there is valid response by the respondents for each cohort member.

Population: All cohort members

Value and Value Labels:

(0) “No score”

Derivation description: Calculated using the mean of the sum of the scores for each question (IBQ1-IBQ14) (only including the questions for which they entered a response code 1-7 (excluding missing or code 8/NA).) So, if only 10 of these questions had a response code 1-7, it would be the sum divided by 10.

dvkessler - “DV Kessler K6 Mental health scales”

Description: Sum of each score for each item in Kessler mental health scale.

Population: All respondents

Value and Value Labels:

(0) “No response”

Derivation description: Calculated by summing the reverse of each score (KES1-KES6) where score is between 0-4, else missing.

dvgad2 - “DV Generalised Anxiety Disorder 2-item”

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

Population: All respondents

Value and Value Labels:

(-8) “Insufficient information”

Derivation description: Calculated from GAD2PHQ2A+GAD2PHQ2B if both questions have valid responses, otherwise the score is -8 due to insufficient information.

dvphq2 - “DV Patient Health Questionnaire 2-item”

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

Population: All respondents

Value and Value Labels:

(-8) “Insufficient information”

Derivation description: Calculated from GAD2PHQ2C+GAD2PHQ2D if both questions have valid responses, otherwise the score is -8 due to insufficient information.

dvpss - "DV Parental Stress Scale"

Description: A score of parental stress related to childcare?
Population: All respondents
Value and Value Labels:

(-8) "Insufficient information"

Derivation description: Six items on this subscale (BON7-BON12) are rated on a 5-point scale, from 1 "Strongly disagree", to 5 "Strongly agree", meaning a sum of scores from 6-30, with higher scores relating to higher levels of stress. The variable is marked as -8 if there are no valid responses for all 6 items.

dvcsi4 - "DV Couple Satisfaction Index (CSI-4)"

Description: Summed score of respondent's relationship satisfaction with partner using CSI-4 scale.
Population: All respondents
Value and Value Labels:

(-8) "Insufficient information"

Derivation description: The CSI used in ELC-FS is the 4-item version, and is asked in each parent's interview (PI, AI, OHP). The first item (RELSAT1) is scored on a 7-point scale, and the remaining 3 items (COUP2, COUP3, COUP4) are scored on a 6-point scale. The total index is scored by summing all scores together if the responses are valid. The variable is marked as -8 if there are no valid responses for all 4 items. The range of scores is from 0-21.

dvssq - "DV Perceived Social Support Questionnaire (FSozUK-6)"

Description: Summed score of each item of the Social Support Questionnaire (6-items)
Population: All respondents

Value and Value Labels:

(-8) “Insufficient information”

Derivation description: The six items (SSQ1-SSQ6) are ranked on a scale from 1-5, and the score across the six can be summed to provide a total estimate of perceived social support. The variable is marked as -8 if there are no valid responses for all 6 items.

A1.7 Household and family variables

dvhhcarers - “DV Type of interviewed parents in household – please note this is unvalidated and may be updated when household grid data is deposited”

Description: CM’s family type per household, according to number and type of parents

Population: All households

Value and Value Labels:

(1) "Both biological parents"

(2) "Biological mother and step-parent"

(3) "Biological mother and partner"

(4) "Biological mother and adoptive parent"

(5) "Biological father and step-parent"

(6) "Biological father and partner"

(7) "Biological father and adoptive parent"

(8) "Biological mother only"

(9) "Biological father only"

Derivation description: Derived from parent relationship to CM (PRELAT_A, PRELAT_AI_A), whether they are a couple/partner (COUPLE_AIWPI), biological relationship to CM (BIOMUM, BIODAD), and gender (PGENDER).

dvnumallhh - “DV Total number of people in household (incl CMs) – please note this is unvalidated and may be updated when household grid data is deposited”

Description: Total number of people in the household, including respondents and CMs.

Population: All households

Derivation description: Summation of the household members from parent response (hhnum), the total respondents per household, and the number of CMs per household (nmultihh).

dvmleave2b_months - “DV Length of parental leave - months”

Description:

Population: All respondents

Value and Value Labels:

(-96) “Missing parental leave date”

Derivation description: Respondents could report how long their parental leave had/will have lasted in either days, weeks or months (LEAVE2BD, LEAVE2BM, LEAVE2BY). This derived variable converts all answers into months.

dvhomelearnscore - “DV Combined home learning score”

Description: It is the sum of frequencies with which parents reported doing each of the home learning activities they were asked about. Higher combined home learning scores indicate a higher frequency of home learning activities.

Population: All respondents

Value and Value Labels:

(-96) “Missing score”

Derivation description: Frequency response options were first transformed into numeric scores (“Never” = 1, "Several times a day" = 6) and then summed across the 12 activities (PLAYF1-PLAYF12). Then the numeric scores are summed.

A1.8 Demographics

dvcmethnic6 - “DV Cohort Member Ethnic Group - 6 category Census class”

Description: Cohort member’s ethnic group as reported by parent respondents, as 6-category census class

Population: All cohort members

Value and Value Labels:

(1) “White”

(2) “Mixed”

(3) “Indian”

(4) “Pakistani and Bangladeshi”

(5) “Black or Black British”

(6) “Other Ethnic group (inc Chinese,Other)”

(-1) “Not applicable”

(-8) “Don't Know”

(-9) “Refusal”

Derivation description: Derived from variable ETHNIC2, or ETHNICw/ETHNICm/ETHNICa/ETHNICb/ETHNICo.

dvcmethnic8 - “DV Cohort Member Ethnic Group - 8 category Census class”

Description: Cohort member’s ethnic group as reported by parent respondents, as 8-category classification

Population: All cohort members

Value and Value Labels:

(1) "White"

(2) "Mixed"

(3) "Indian"

(4) "Pakistani"

(5) "Bangladeshi"

(6) "Black Caribbean"

(7) "Black African"

(8) "Other Ethnic group (inc Chinese, Other)"

(-1) “Not applicable”

(-8) “Don't Know”

(-9) “Refusal”

Derivation description: Derived from variable ETHNIC2, or ETHNICw/ETHNICm/ETHNICa/ETHNICb/ETHNICo.

dvcmethnic11 - “DV Cohort Member Ethnic Group - 11 category Census class”

Description: Cohort member’s ethnic group as reported by parent respondents, as 11-category census class

Population: All cohort members

Value and Value Labels:

(1) "White"

(2) "Mixed"

(3) "Indian"

(4) "Pakistani"

(5) "Bangladeshi"

(6) "Other Asian"

(7) "Black Caribbean"

(8) "Black African"

(9) "Other Black"

(10) "Chinese"

(11) "Other Ethnic group"

(-1) “Not applicable”

(-8) “Don't Know”

(-9) “Refusal”

Derivation description: Derived from variable ETHNIC2, or ETHNICw/ETHNICm/ETHNICa/ETHNICb/ETHNICo.

dvethnic4 - “DV Respondent ethnicity - 4 categories”

Description: Respondent’s self-reported ethnic group, classified into 4 categories

Population: All respondents

Value and Value Labels:

(1) “White”

(2) “Asian or Asian British”

(3) “Black or Black British”

(4) “Mixed or Other ethnic groups”

(-1) “Not applicable”

(-8) “Don’t know”

(-9) “Refusal”

Derivation description: Derived from PETHNIC_A.

dvethnic6 - “DV Respondent ethnicity - 6 category Census class”

Description: Respondent’s self-reported ethnic group, classified into 6 category Census class

Population: All respondents

Value and Value Labels:

(1) “White”

(2) “Mixed”

(3) “Indian”

(4) “Pakistani and Bangladeshi”

(5) “Black or Black British”

(6) “Other Ethnic group (inc Chinese,Other)”

(-1) “Not applicable”

(-8) “Don’t know”

(-9) “Refusal”

Derivation description: Derived from PETHNIC_A.

dvethnic8 - “DV Respondent ethnicity - 8 category Census class”

Description: Respondent’s self-reported ethnic group, classified into 8 category Census class.

Population: All respondents

Value and Value Labels:

(1) “White”

(2) “Mixed”

(3) “Indian”

(4) “Pakistani”

(5) “Bangladeshi”

(6) “Black Caribbean”

(7) “Black British”

(8) “Other Ethnic group (inc Chinese,Other)”

(-1) “Not applicable”

(-8) “Don’t know”

(-9) “Refusal”

Derivation description: Derived from PETHNIC_A.

dvethnic11 - “DV Respondent ethnicity - 11 category Census class”

Description: Respondent’s self-reported ethnic group, classified into 11 category Census class.

Population: All respondents

Value and Value Labels:

(1) “White”

(2) “Mixed”

(3) “Indian”

(4) “Pakistani”

(5) “Bangladeshi”

(6) “Other Asian”

(7) “Black Caribbean”

(8) “Black British”

(9) “Other Black”

(10) “Chinese”

(11) “Other Ethnic group”

(-1) “Not applicable”

(-8) “Don’t know”

(-9) “Refusal”

Derivation description: Derived from PETHNIC_A.

dvethnic4 - “DV Partner’s ethnicity - 4 categories”

Description: Partner’s self-reported ethnic group, classified into 4 categories

Population: PI and OHP respondents with a co-resident partner

Value and Value Labels:

(1) “White”

(2) “Asian or Asian British”

(3) “Black or Black British”

(4) “Mixed or Other ethnic groups”

(-1) “Not applicable”

(-8) “Don’t know”

(-9) “Refusal”

Derivation description: Derived from ETHNICP_A

dvethnic6 - “DV Partner’s ethnicity - 6 category Census class”

Description: Partner’s self-reported ethnic group, classified into 6 category Census class

Population: PI and OHP respondents with a co-resident partner

Value and Value Labels:

(1) “White”

(2) “Mixed”

(3) “Indian”

(4) “Pakistani and Bangladeshi”

(5) “Black or Black British”

(6) “Other Ethnic group (inc Chinese,Other)”

(-1) “Not applicable”

(-8) “Don’t know”

(-9) “Refusal”

Derivation description: Derived from ETHNICP_A

dvethnic8 - “DV Partner’s ethnicity - 8 category Census class”

Description: Partner’s self-reported ethnic group, classified into 8 category Census class.

Population: PI and OHP respondents with a co-resident partner

Value and Value Labels:

(1) “White”

(2) “Mixed”

(3) “Indian”

(4) “Pakistani”

(5) “Bangladeshi”

(6) “Black Caribbean”

(7) “Black British”

(8) “Other Ethnic group (inc Chinese,Other)”

(-1) “Not applicable”

(-8) “Don’t know”

(-9) “Refusal”

Derivation description: Derived from ETHNICP_A

dvethnic11 - “DV Partner’s ethnicity - 11 category Census class”

Description: Partner’s self-reported ethnic group, classified into 11 category Census class.

Population: PI and OHP respondents with a co-resident partner

Value and Value Labels:

(1) “White”

(2) “Mixed”

(3) “Indian”

(4) “Pakistani”

(5) “Bangladeshi”

(6) “Other Asian”

(7) “Black Caribbean”

(8) “Black British”

(9) “Other Black”

(10) “Chinese”

(11) “Other Ethnic group”

(-1) “Not applicable”

(-8) “Don’t know”

(-9) “Refusal”

Derivation description: Derived from ETHNICP_A

dvage_at_cmbirth - “DV Respondent age at birth of CM”

Population: All respondents

Value and Value Labels:

(-1) “Not applicable”

Derivation description: Calculated by finding the difference between respondent’s age (P_AGE or derived from PDOB or PDOBD/PDOBM/PDOBY) and Cohort Member’s date of birth (CDOB or CDOBD/CDOBM/CDOBY).

dvage_at_cmbirth_grouped - “DV Respondent age at birth of CM - grouped”

Population: All respondents

Value and Value Labels:

(1) “19 and below”

(2) “Between 20 to 29”

(3) “Between 30 to 39”

(4) “40 and above”

(-1) “Not applicable”

Derivation description: Grouping variable dvage_at_cmbirth into age groups, to be consistent with MCS1.

dv_prelat_tocm - “DV Respondent relationship to CM”

Description: Relationship of all parent respondents to the CM.

Population: All respondents

Value and Value Labels:

(1) “Biological parent”

(2) “Adoptive parent”

(3) “Foster parent”

(4) “Special guardian / kinship carer”

(5) “Step parent”

(6) “Something else”

Derivation description: Combined from PRELAT_A or PRELAT_AI_A, and BIOMUM/BIODAD.

A1.8 Relationship & Partnership Status

dvhhrelstatus - “DV Relationship between Parents/Carers in the household – please note this is unvalidated and may be updated when household grid data is deposited”

Description: Type of couple relationship between parents/carers in the household

Population: All households

Value and Value Labels:

(1) “Married”

(2) “Cohabiting”

(3) “Neither”

(-1) “Not applicable”

Derivation description: Calculated if HPARTNER=1 and HBIOPMARR=1 then 1, or HBIOPMARR=2 then 2, else 3. If HPARTNER!=1 then -1.
Whether there is a partner in the household is derived from HPARTNER, and HBIOPMARR find the relationship between these people in the household.

dvinrelat - “DV Whether PI in a relationship with cohabiting partner or a non-resident parent – please note this is unvalidated and may be updated when household grid data is deposited”

Description: Whether PI is in a relationship with either a cohabiting partner or a non-resident parent.

Population: All households

Value and Value Labels:

(-1) “Not applicable”

(1) “Yes”

(2) “No”

Derivation description: This variable flags whether the PI report any members of their household as their partner or spouse or report that they consider themselves to be “in a relationship” with the cohort child/children’s other biological parent living outside of the household, using HPARTNER/OHPREL. Note that this does not include any relationships of this kind with people who neither live in the household nor are the cohort child/children’s biological parent.

Appendix 2: Non-response weight derivation: Response models

Table A2.1. Response model for England.

Variable	Odds ratio	95% CI
Child sex (SF_CHILDSEX)		
Female	1.00	(ref)
Male	0.99	0.95, 1.03
Child ethnicity (BABY_ETHNICITY)		
Asian or Asian British – Any other Asian background	1.00	(ref)
Asian or Asian British – Bangladeshi	1.07	0.91, 1.25
Asian or Asian British – Indian	0.99	0.48, 2.06
Asian or Asian British – Pakistani	1.29	1.06, 1.58
Black or Black British – African	0.93	0.46, 1.86
Black or Black British – Any other Black background	1.23	0.57, 2.63
Black or Black British – Caribbean	0.99	0.72, 1.35
Mixed – Any other mixed background	1.37	0.63, 2.97
Mixed – White and Asian	1.08	0.92, 1.27
Mixed – White and Black African	1.39	1.07, 1.81
Mixed – White and Black Caribbean	1.42	1.03, 1.96
Not Stated	1.13	0.89, 1.42
Other Ethnic Groups – Any other ethnic group	1.05	0.87, 1.27
White – Any other White background	1.12	0.98, 1.28
White – British	1.17	1.02, 1.35
White – Irish	1.02	0.68, 1.55
Child birth weight (BIRTH_WEIGHT)		
<2500g	1.00	(ref)
2500-2999g	1.03	0.94, 1.13
3000-3499g	0.99	0.90, 1.09
3500-3999g	1.04	0.94, 1.15
4000+g	1.08	0.95, 1.22

Variable	Odds ratio	95% CI
Child gestation (GESTATION)		
<37 weeks	1.00	(ref)
37 weeks	0.91	0.83, 1.00
38 weeks	0.98	0.89, 1.08
39 weeks	0.97	0.89, 1.07
40 weeks	0.96	0.87, 1.06
41+ weeks	0.98	0.88, 1.10
Mother's age at birth (MOTHER_AGEATBIRTH)		
<20	1.00	(ref)
20-24	1.04	0.90, 1.21
25-29	1.05	0.90, 1.23
30-34	1.02	0.87, 1.20
35-39	1.08	0.91, 1.28
40+	1.14	0.94, 1.38
Father's age at birth (FATHER_AGEATBIRTH)		
<20	1.00	(ref)
20-24	1.06	0.86, 1.30
25-29	1.01	0.81, 1.25
30-34	1.05	0.84, 1.31
35-39	1.13	0.90, 1.42
40+	1.12	0.89, 1.41
Mother's country of birth (MOTHER_CCPOB)		
Any other country	1.00	(ref)
Bangladesh	1.06	0.85, 1.33
England	0.91	0.84, 0.99
India	0.91	0.69, 1.21
Nigeria	1.11	0.90, 1.36
Pakistan	1,08	0.94, 1.24
Romania	0.96	0.74, 1.25
Wales	0.82	0.62, 1.09

Variable	Odds ratio	95% CI
Father's country of birth (FATHER_CCPOB)		
Any other country	1.00	(ref)
Bangladesh	1.14	0.91, 1.42
England	1.08	1.00, 1.18
India	0.95	0.74, 1.24
Nigeria	0,93	0.76, 1.13
Pakistan	1.04	0.91, 1.19
Romania	1.04	0.81, 1.33
Wales	1.07	0.71, 1.63
IMD decile (imd_decile)		
1	1.00	(ref)
2	0.93	0.87, 0.99
3	1.00	0.92, 1.08
4	0.97	0.88, 1.07
5	0.99	0.89, 1.09
6	0.91	0.81, 1.01
7	1,01	0.89, 1.14
8	1.01	0.90, 1.14
9	1.06	0.94, 1.20
10	0.94	0.82, 1.08
Incentive group		
£5 unconditional & £10 conditional	1.00	(ref)
£5 unconditional & £20 conditional	1.04	0.97, 1.12
Bib & £10 conditional	0.91	0.85, 0.97
Bib & £20 conditional	1.00	0.94, 1.08
No unconditional & £10 conditional	0.97	0.91, 1.05
No unconditional & £20 conditional	0.99	0.92, 1.06
Birth in marriage (BIRTH_INMARRIAGE)		
Birth in marriage	1.00	(ref)
Birth outside of marriage	0.85	0.73, 0.98

Variable	Odds ratio	95% CI
Birth informant (BIRTH_INFORMANTQUAL)		
Father/Parent	1.00	(ref)
Father/Parent and Mother (Joint Informants)	1.12	0.96, 1.31
Mother	1.02	0.96, 1.09
Region name (RegionName)		
East Midlands	1.00	(ref)
East of England	0.95	0.84, 1.08
London	0.87	0.78, 0.97
North East	0.98	0.85, 1.12
North West	0.92	0.82, 1.03
South East	0.94	0.83, 1.06
South West	0.91	0.79, 1.05
West Midlands	0.95	0.85, 1.06
Yorkshire and The Humber	0.98	0.87, 1.10
Ethnicity/area deprivation group (group)		
Black African or Caribbean ethnicity	1.00	(ref)
Pakistani or Bangladeshi ethnicity	0.97	0.47, 1.83
Living in a deprived PSU and **not** in one of the above ethnic groups	0.92	0.46, 1.85
Not living in a deprived PSU and **not** in one of the above ethnic groups	0.93	0.46, 1.85
Stratum (stratum)		
All other PSUs in deprived areas (defined as more than half in a deprived LSOA)	1.00	(ref)
All other PSUs not previously allocated to a stratum	1.01	0.92, 1.10
PSUs with seven or more Bangladeshi and Pakistani children	1.06	0.98, 1.14
PSUs with seven or more Black African and Black Caribbean children	1.01	0.93, 1.09
PSUs with three or more Black African, Black Caribbean, Bangladeshi and Pakistani children	0.95	0.88, 1.02

Table A2.2. Response model for Wales.

Variable	Odds ratio	95% CI
Child sex		
Female	1.00	(ref)
Male	1.06	0.4, 2.33
Child ethnicity		
Any other ethnic group	1.00	(ref)
Asian or Asian British - Any Asian background	0.52	0.26, 1.05
Black or Black British - Any Black background	1.70	0.75, 3.86
Mixed Background	0.78	0.46, 1.33
Not Stated	0.71	0.47, 1.09
White - Any other White background	1.04	0.53, 2.04
White - British	0.78	0.51, 1.20
Child birth weight		
<2500g	1.00	(ref)
2500-2999g	1.05	0.87, 1.27
3000-3499g	1.01	0.83, 1.22
3500-3999g	1.07	0.87, 1.32
4000+g	1.12	0.89, 1.40
Child gestation		
<37 weeks	1.00	(ref)
37 weeks	0.95	0.78, 1.16
38 weeks	0.97	0.80, 1.18
39 weeks	0.87	0.72, 1.06
40 weeks	0.90	0.74, 1.09
41+ weeks	0.92	0.75, 1.13
Mother’s age at birth		
<20	1.00	(ref)
20-24	0.97	0.73, 1.28
25-29	1.13	0.83, 1.53
30-34	1.13	0.82, 1.55

Variable	Odds ratio	95% CI
35-39	1.06	0.76, 1.48
40+	1.21	0.81, 1.81
Father's age at birth		
<20	1.00	(ref)
20-24	0.99	0.69, 1.43
25-29	0.86	0.59, 1.27
30-34	1.01	0.69, 1.49
35-39	1.00	0.68, 1.49
40+	0.99	0.66, 1.49
Mother's country of birth		
Any other country	1.00	(ref)
Bangladesh	0.74	0.21, 2.56
England	1.13	0.92, 1.38
India	3.05	0.93, 10.08
Nigeria	0.99	0.30, 3.26
Pakistan	0.55	0.25, 1.21
Romania	0.60	0.27, 1.34
Wales	1.00	0.83, 1.20
Father's country of birth		
Any other country	1.00	(ref)
England	1.26	1.02, 1.56
India	0.77	0.29, 2.03
Nigeria	0.77	0.30, 1.99
Pakistan	1.99	0.74, 5.37
Romania	1.99	0.88, 4.48
Wales	1.23	1.01, 1.50
IMD decile		
1	1.00	(ref)
2	0.92	0.79, 1.08

Variable	Odds ratio	95% CI
3	0.76	0.65, 0.90
4	0.84	0.71, 0.99
5	1.00	0.84, 1.18
6	0.98	0.83, 1.16
7	1.01	0.84, 1.22
8	1.09	0.92, 1.29
9	0.91	0.75, 1.10
10	0.97	0.82, 1.16
Incentive group		
£5 unconditional & £10 conditional	1.00	(ref)
£5 unconditional & £20 conditional	1.06	0.92, 1.21
Bib & £10 conditional	0.92	0.81, 1.06
Bib & £20 conditional	0.92	0.80, 1.05
No unconditional & £10 conditional	0.89	0.78, 1.02
No unconditional & £20 conditional	0.88	0.77, 1.01
Birth in marriage		
Birth in marriage	1.00	(ref)
Birth outside of marriage	0.87	0.70, 1.08
Birth informant		
Father/Parent	1.00	(ref)
Father/Parent and Mother (Joint Informants)	1.08	0.85, 1.37
Mother	0.96	0.84, 1.09

Table A2.3. Response model for Scotland.

Variable	Odds ratio	95% CI
Child sex (SF_CHILDSEX)		
Female	1.00	(ref)
Male	0.96	0.88, 1.04
Child ethnicity (BABY_ETHNICITY)		
Asian or Asian British - Any other Asian background	1.00	(ref)
Any mixed or multiple ethnic groups	0.63	0.35, 1.11
Arab, Arab Scottish or Arab British	0.34	0.16, 0.75
Bangladeshi, Bangladeshi Scottish or Bangladeshi British	0.57	0.25, 1.31
Chinese, Chinese Scottish or Chinese British	0.96	0.32, 2.86
Indian, Indian Scottish or Indian British	0.49	0.25, 0.98
Not known	0.54	0.31, 0.95
Other African	0.92	0.30, 2.81
Other Asian, Asian Scottish, Asian British	0.41	0.19, 0.89
Other ethnic group	0.64	0.21, 1.93
Other white ethnic group	0.62	0.36, 1.09
Pakistani, Pakistani Scottish or Pakistani British	0.60	0.34, 1.06
Polish	0.45	0.23, 0.90
White Irish	0.32	0.13, 0.81,
White Other British	0.52	0.30, 0.91
White Scottish	0.61	0.36, 1.03
Child birth weight (BIRTH_WEIGHT)		
<2500g	1.00	(ref)
2500-2999g	0.89	0.70, 1.12
3000-3499g	0.91	0.73, 1.15
3500-3999g	0.86	0.68, 1.10
4000+g	0.92	0.71, 1.20
Child gestation (GESTATION)		
<37 weeks	1.00	(ref)
37 weeks	1.06	0.84, 1.33

Variable	Odds ratio	95% CI
38 weeks	1.00	0.78, 1.27
39 weeks	1.06	0.84, 1.33
40 weeks	0.98	0.78, 1.25
41+ weeks	1.02	0.80, 1.29
Mother's age at birth (MOTHER_AGEATBIRTH)		
<20	1.00	(ref)
20-24	0.79	0.47, 1.35
25-29	0.82	0.47, 1.42
30-34	0.78	0.45, 1.37
35-39	0.76	0.43, 1.34
40 - 44	0.75	0.41, 1.37
45+	1.36	0.53, 3.50
Father's age at birth (FATHER_AGEATBIRTH)		
<20	1.00	(ref)
20-24	1.18	0.67, 2.07
25-29	1.23	0.68, 2.21
30-34	1.33	0.73, 2.42
35-39	1.33	0.72, 2.43
40-44	1.25	0.67, 2.30
45+	1.26	0.67, 2.36
IMD decile (imd_decile)		
1	1.00	(ref)
2	0.95	0.78, 1.15
3	1.03	0.86, 1.23
4	1.00	0.82, 1.23
5	1.03	0.85, 1.27
6	1.05	0.85, 1.30
7	0.96	0.80, 1.16
8	1.15	0.94, 1.39
9	0.92	0.76, 1.11
10	1.04	0.85, 1.27

Variable	Odds ratio	95% CI
Incentive group		
£5 unconditional & £10 conditional	1.00	(ref)
£5 unconditional & £20 conditional	0.97	0.84, 1.12
Bib & £10 conditional	0.89	0.77, 1.03
Bib & £20 conditional	1.02	0.88, 1.18
No unconditional & £10 conditional	0.95	0.82, 1.11
No unconditional & £20 conditional	0.93	0.80, 1.08
Father's Social Grade (fsclacd)		
Higher professional occupations	1.00	(ref)
Intermediate occupations	0.96	0.80, 1.16
Large employers and higher managerial occupations	0.92	0.66, 1.28
Lower managerial and professional occupations	1.08	0.93, 1.25
Lower supervisory and technical occupations	1.08	0.91, 1.29
Routine occupations	0.96	0.80, 1.15
Semi-routine occupations	1.02	0.84, 1.23
Small employers and own account workers	0.86	0.71, 1.04
Students, not stated or not classifiable	0.89	0.69, 1.16
Mother's Social Grade		
Higher Professional Occupations	1.00	(ref)
Intermediate occupations	0.91	0.77, 1.08
Large employers and higher managerial occupations	0.84	0.57, 1.24
Lower managerial and professional occupations	0.97	0.83, 1.14
Lower supervisory and technical occupations	0,85	0.60, 1.21
Never worked and long term unemployed	1.05	0.61, 1.80
Routine occupations	1.11	0.88, 1.39
Semi-routine occupations	0.86	0.72, 1.03
Small employers and own account workers	0.74	0.58, 0.94
Students, not stated or not classifiable	0.84	0.69, 1.02
Mother's Ethnicity (mother_ethnic_name)		
Chinese, Chinese Scottish or Chinese British	1.00	(ref)

Variable	Odds ratio	95% CI
Any mixed or multiple ethnic groups	1.24	0.61, 2.53
Any other white ethnic group	1.92	1.04, 3.56
Arab, Arab Scottish or Arab British	2.78	0.95, 8.17
Bangladeshi, Bangladeshi Scottish or Bangladeshi British	2.21	0.87, 5.61
Caribbean, Caribbean Scottish or Caribbean British	0.89	0.27, 2.92
Indian, Indian Scottish or Indian British	1.59	0.65, 3.90
Not Known	1.47	0.85, 2.54
Other African	1.47	0.74, 2.91
Other Asian, Asian Scottish or Asian British	1.85	0.90, 3.80
Other ethnic group	1.35	0.66, 2.78
Pakistani, Pakistani Scottish or Pakistani British	1.88	1.00, 3.56
Refused/Not provided	1.46	0.77, 2.77
White Irish	2.26	1.06, 4.83
White Other British	1.61	0.91, 2.85
White Polish	1.86	0.89, 3.86
White Scottish	1.53	0.88, 2.67
Parity(Parity)		
0	1.00	(ref)
1+	1.04	0.97, 1.12