Data Resource Profile: ECHILD only-children and siblings (ECHILD-oCSib): a national cohort of linked health, education and social care data on mothers and children in England

Qi Feng[1,*], Georgina Ireland[1], Ruth Gilbert[1], and Katie Harron[1]

Abstract

Introduction
Sibling dynamics play a crucial role in individual development, health and wellbeing. We established a national birth cohort using administrative health, education and social care data in England featuring clusters of mothers and their children (mothers and only-children, MoC; and mothers and siblings, MSib).

Methods
From 13.6 million mother-baby pairs from births between April 1997 and January 2022 captured in Hospital Episode Statistics in England, we identified MoC and MSib clusters by identifying livebirths linked to the same mother. We compared only-children and children with siblings, by ethnicity, sociodemographic variables, and birth characteristics. We calculated birth intervals for children with siblings.

Results
We identified 4,086,648 MoC and 3,957,856 MSib clusters. Compared with only-children, children with siblings were more likely to be Asian, live in more deprived areas, and have younger mothers, but were less likely to be overdue births ($>=42$ weeks), or to have very low birth weight (<1500g). Children with siblings were also less likely to have been admitted to special neonatal care after birth compared to only-children. Among the MSib clusters, sibship sizes varied between 2 and 15, with a mean of 2.4 children per mother. The median birth interval was 3.0 years.

Conclusion
This national cohort ECHILD-oCSib of 4.1 million MoC and 4.0 million MSib clusters in England is an important resource for investigating the effects of maternal exposures, sibling dynamics and their interplay on individual development, health and wellbeing. Potential sources of bias should be considered in analyses of these data.

Keywords
only child; siblings; data linkage; hospital records; Hospital Episode Statistics; National Pupil Database

Key features

- We derived a national cohort of 4.1 million clusters of mothers and only-children and 4.0 million clusters of mothers and siblings using administrative health, education and social care data in England.

- Compared with only-children, children with siblings were more likely to be Asian, live in more deprived areas, and have younger mothers, but less likely to be overdue births ($>=42$ weeks of gestation), or to have very low birth weight (<1500 g).

- Among children with siblings, sibship sizes varied between 2 and 15, with a mean of 2.4 children per mother (median $=2.0$). The median birth interval was 3.0 years.

- The cohort is linked to longitudinal administrative data on health, education and social care use, and provides a valuable opportunity to investigate the effects of maternal factors, sibling dynamics, and their interaction on children development, health, education and wellbeing.

- The data can be accessed as part of ECHILD database.

Background

In the UK, approximately 57% of families comprise two or more dependent children, and more than half of the population grows up with sibling(s) [1]. Previous research has revealed notable variation in the development, health and overall wellbeing in various stages of the life course between children with sibling(s) and only-children (i.e., those without sibling(s)). Siblings, through their interactions, contribute to the development of health behaviours, social support systems, personalities, and wellbeing [2]. Analysis of Swedish National Registry data has shown that compared to individuals raised with siblings, only-children have lower stature and fitness levels, and are more likely to be obese in late adolescence and experience higher mortality in later life [3]. Systematic review evidence has highlighted that children with siblings tended to engage in higher levels of physical activity relative to their only-children counterparts [4]. Other studies suggest that only-children and children with siblings differ in adulthood obesity [5, 6], metabolism [7], life satisfaction, violent and altruistic behaviour [8], academic performance and psychological adjustment [9].

Sibship characteristics, encompassing factors such as sibship size, birth order, age difference, and sibling relationship dynamics, also significantly influence individual outcomes of development, education, health and wellbeing [10–15]. For example, a British cohort found that the presence of older siblings was associated with relatively better mental health, while presence of younger siblings was associated with poorer mental health [16]. A Swedish study also showed that almost half of the variation in school attainment was attributable to family-level factors, including sibship size, birth spacing and sibling sex composition [17]. A study of 26 countries similarly found a link between sibship size and individual educational attainment, and observed its variation across regions [18]. Living with a chronically ill sibling was found to increase the risk of mental, behavioural and social problems [19–21].

Numerous sibling cohorts have been established around the world, with a predominant focus on twin pairs, such as the US National Heart, Lung, and Blood Institute Male Veteran Twin Study [22], Swedish Twin Registry [23], Danish Twin Registry [24], Norwegian Twin Registry [25], Chinese National Twin Registry [26], alongside large collaborative consortia, for example, CODATwin study [27]. While twin studies provide distinctive insights in epidemiological research, twin births account for only a small proportion of all births, for instance, 1.5% in the UK [28]. It remains unclear whether the different characteristics of twin and non-twin siblings might undermine the generalisability of findings from twin-based studies to broader sibling population [29–32].

Beyond twin-based approaches, sibling cohorts encompassing a wider spectrum of sibling types offer larger coverage of general population and valuable perspectives. Examples include the California sibling study [33], the Swedish National Registry-based sibling study [34], and the Danish sibling study [35]. These cohorts serve as crucial complements to twin cohorts, yet they often lack comprehensive information on maternal factors. A UK-based sibling cohort integrated data from four birth cohorts, and identified siblings through their residence information [36]. However, this compilation comprised survey cohorts with a modest sample size of 41,000 children, resulting in limited representativeness. Therefore, there is an imperative to establish a large sibling cohort that has comprehensive coverage of population, incorporates multiple data sources on mothers and children, and facilitates extensive follow-up. The establishment of such a sibling cohort presents an important resource for understanding the effects of and the interplay between maternal factors, and sibling-related dynamics in shaping child outcomes.

In this paper, we first provide a comprehensive overview of the background, data sources and methods for establishing a national cohort of clusters of mothers and their child(ren) in England. In this cohort, we have special focus on clusters of mothers and only-children (MoC) and clusters of mothers and siblings (MSib). We describe the basic characteristics of the cohorts in the Results, and discuss the strengths, limitations and their potential use in child research in the Discussion. Lastly, we discuss how to access the data.

Methods

Data sources and participants

We used the Education and Child Health Insights from Linked Data (ECHILD) [37] and ECHILD mother-baby (ECHILD-MB) cohort [38] to establish the MoC and MSib cohorts.

ECHILD is a comprehensive national linkage of administrative health and education databases for children and young people born after 1984 in England. The linkage involved the integration of administrative hospital data from the National Health Service (NHS) Hospital Episode Statistics (HES) [39] and education and children's social care data from the Department for Education (DfE) National Pupil Database (NPD) [40].

HES includes records of all hospital activities funded by NHS in England, including deliveries, births, inpatient admissions, outpatient appointments, accident and emergency attendances, mortality, demographics and standardised codes for diagnoses and procedures [39].

NPD contains records related to state-funded educations and children's social care services, incorporating several data modules that are collected by DfE. NPD includes information from different educational settings about pupil's characteristics, including age, gender, ethnicity, special educational needs and free school meals. Educational outcomes are captured, including absences, exclusions, attainment in national assessments and examination, and participation in post-16 education. NPD also features two social care modules, including Children in Need (for children referred to social care services) and Children Looked after Return (for children in care, referred to as looked after children in the UK) [40].

ECHILD ensures the anonymity of individuals by removing all direct identifiers. The linkage between HES and NPD was performed by NHS England. Briefly, DfE securely transferred identifiers from NPD to NHS England, and NHS England performed deterministic linkage to link direct identifiers to create an anonymized linkage spine connecting the NPD and HES [41, 42]. Table 1 shows the key datasets in ECHILD and other cohorts derived from it.

Table 1: Key data sources included in ECHILD-oCSib cohort

Data source	Description	Year coverage	Age coverage (years)	Key variables
HES				
Admitted patient care	Diagnoses, operations, operation dates, consultant specialty	1997-2022	All	Diagnoses, operations, operation dates, etc.
Critical care	Critical care start and end dates, number of days of support by organ group, discharge destination	2006-2022	Adults	Diagnosis codes, etc.
Accident and emergency / Emergency Care Services Dataset	Type of attendance, mode of arrival, treatments, duration	2006-2022	All	Diagnosis codes, etc.
Outpatient	Type of appointment, outcome of appointment, medical staff type seeing patient, duration of elective wait	2002- 2022	All	Diagnosis codes, etc.
ONS linked mortality death registration	Month and year of death, underlying cause of death	1997- 2022	All	Date of death, cause of death, etc.
Birth notification	Birth notification	2001-2022	Birth	Birth weight, gestational age
Birth registration	Birth registration	1996-2022	Birth	Sex, multiple indicator, parents' country of birth and occupation
NPD				
Early Years Census	All 2 to 4-year olds in state-funded early years care and education	2007-2022	2-4	Age, gender, ethnicity, SEN
School census pupil level	All pupils in state-maintained educational settings, excluding hospital schools	2005-2022	2-16	Age; gender; ethnicity; SEN; FSM eligibility; language
Pupil referral Unit census	All pupils in a PRUs (non-mainstream schools maintained by the state)	2009-2013	2-16	Age; gender; ethnicity; SEN; FSM eligibility; language
Alternative provision census	All pupils in non-mainstream, non-maintained educational settings for whom the state are covering tuition costs	2007-2022	2-16	Age; gender; ethnicity; SEN; FSM eligibility
Absences	All pupils in state-maintained educational settings, excluding boarding pupils	2005-2022	4-16	Number of absences; numbers that were authorized and unauthorized
Exclusions	All pupils in state-maintained educational settings	2001-2021	2-16	Number of fixed period exclusions; number of permanent exclusions
Early years foundation stage profile	All children at the end of the Early Years Foundation Stage of education	2002-2019	3-5	Early years practitioner assessment scores
KS1 assessment	All children at the end of KS1	1997-2022	5-7	Teacher assessments scores
KS2 assessment	All children at the end of KS2	1995-2022	7-11	Teacher assessments scores
KS3 assessment	All children at the end of KS3	1998-2013	11-14	Teacher assessments scores
KS4 qualification	All pupils in KS4, including those in private schools	2001-2021	14-16	Entry for and attainment in GCSE and equivalent qualifications
KS5 qualification	All pupils in KS5, including those in private schools	2002-2021	16-18	Entry for and attainment in A-level and equivalent qualifications
National Client caseload information system	All young people aged 16–25 who have an SEN or disability	2010-2022	16-25	Post-16 activity; not in education, employment or training indicator
Children in Need Census	Referrals to children's social care and all children in need	2008-2022	2-16	Referral date; category of need; start date of child protection plan
Children Looked After Return	All children who are looked after	1991-2021	2-16	Placement start and end date; type of placement setting; legal basis for placement

Notes: HES = Hospital Episode Statistics; ONS = Office for National Statistics; N/A, not available.

Information on diagnoses, treatments and procedures for each episode of care is recorded by clinical coders based on patient care records and/or discharge summaries using standardized codes. In the Admitted Patient Care, Critical Care and Outpatient modules, diagnoses are recorded using the International Classification of Disease (ICD) version 10, and treatments and procedures are recorded using the Office of Population Censuses and Surveys (OPCS) version 4. In the Accident and Emergency module, bespoke codes are used to record diagnoses and treatments14; however, these are much more limited than ICD-10 and OPCS-4 codes.

NPD, National Pupil Database; PRU, pupil referral unit; SEN, special educational needs; FSM, free school meals; GCSE, General Certificate of Secondary Education (national examinations taken by students at the end of compulsory education); KS, Key Stage. The School Census Pupil Level module is collected on a termly basis in October (Autumn census), January (Spring census) and May (Summer census). The other education census modules are collected in January only.

The ECHILD-MB cohort, nested within ECHILD, specifically focuses on mother-baby pairs. In HES, delivery records and birth records are separately documented for mothers and babies, with no routine process for identifying mother-baby pairs for NHS England. We employed a validated linkage algorithm, encompassing both deterministic and probabilistic linkage methods, using information of demographics, geographic, delivery and birth characteristics to identify these pairs [43]. The algorithm successfully linked 13.6 million out of 14.5 million (overall linkage rate 94.1%) birth records to maternal delivery records between April 1997 and January 2022. External validation of the mother-baby linking status against an independent administrative dataset (Community Service Data Set) covering a subset of the ECHILD-MB cohort demonstrated low levels of linkage error. ECHILD-MB covers 87% of all livebirths in England and is representative of national birth statistics [38]. We excluded stillbirths from the ECHILD-MB cohort before deriving the MoC and MSib cohorts. (Supplementary Table 1).

Statistical analysis

Based on the number of livebirths identified for each mother in HES, we identified clusters of MoC and of MSib to form the ECHILD-MoC cohort and the ECHILD-MSib cohort. In ECHILD-MSib cohort, we integrated information across births from a multiple pregnancy (such as twins, triples, etc.) to impute missing data for gestational age and maternal age. We calculated the sibship size as the total number of livebirths (not deliveries) up to January 2022.

We compared the basic characteristics between the MoC and MSib cohorts, as well as by sibship size in the MSib cohort, including sex, birth weight (g), gestational age (weeks), mother's ethnic background (White, Asian, Black, Mixed, and Others), maternal age (years), and mother's Index of Multiple Deprivation (IMD) quintiles (based on residential postcode). Differences between groups were compared using standardized differences and their 95% confidence intervals. When calculating the summary statistics of these characteristics for each group of sibship size, we used individual-level data, instead of cluster-level data. For maternal characteristics (maternal age, ethnic background, IMD), the same mother may be included multiple times if she had two or more children. Based on whether a variable was time varying or not, there were three categories: (1) time-constant variables, such as maternal ethnicity; (2) delivery-specific variables, such as gestational age, maternal age, and maternal IMD; and (3) baby-specific variables, such as birth weight, baby sex, neonatal care admission, and birth status.

In the MSib cohort, we assessed the proportion of boys and of multiple births overall and by sibship size, where the latter was calculated as the number of children who were born in multiple births divided by the total number of children. We also compared the characteristics of children based on their birth order (first born vs. later born children), after removing the records where the first birth was a multiple birth.

We calculated birth interval and inter-pregnancy interval. Birth interval was defined as the time period between a delivery and the next delivery; interpregnancy-interval was defined as the time period between delivery and estimated conception date of the subsequent pregnancy. The conception date was estimated by subtracting gestational age from the derived delivery date. Derived delivery date was estimated from mothers' delivery records using the same algorithm as used in a previous study [44]. When gestational age was missing (for 17.8% of mother-baby dyads), we used the median value (39.0 weeks). By definition, birth interval and inter-pregnancy interval for the first delivery was zero. We summarised the mean, median and interquartile rage of birth interval and interpregnancy interval, overall and by sibship size.

Results

In the ECHILD-MB cohort, after excluding 29 440 stillbirths, we identified a total of 8 044 504 mothers with at least one livebirth. Among these, 4 086 648 (50.8%) had a single livebirth, constituting the MoC cohort, while the remaining 3 957 856 (49.2%) had more than one livebirth, forming the MSib cohort (Figure 1).

Compared to the only-children, children with siblings exhibited different characteristics. They were more likely to be Asian (12.0% vs. 10.3%) and less likely to have a Mixed ethnicity background (4.4% vs. 5.3%). Children with siblings were also more likely to live in the most deprived quintile (28.5% vs. 24.4%), to have younger mothers (29.0 vs. 30.0 years), less likely to be overdue births (>=42 weeks; 3.2% vs. 3.8%), have higher birth weight (3263.6 vs. 3252.0g), and less likely to have very low birth weight (>1500g; 0.8% vs. 1.0%). They were also less likely to be admitted to special neonatal care following birth (12.5% vs. 13.8%) (Table 2).

Within the MSib cohort, the number of children in each cluster varied between 2 and 15, with a mean of 2.4 children per mother (median 2.0, mode 2.0, interquartile range 2.0–3.0). Of mothers with two or more children, 71.1% (n = 2 815 657) had two children, 20.8% (n = 821 999) had three children, 5.7% (n = 226 515) had four children, and 2.4% had more than four children.

Among the sibling clusters with two children, 51.4% consisted of one boy and one girl, 23.1% two boys, and 25.5% two girls. Distributions of sex for higher order births are provided in Supplementary Table 2. The percentage of sibling clusters that included one or more pregnancy with multiple births ranged from 2.2% in 2-sibling-clusters to 12.1% in clusters with more than five children (Figure 2).

Compared to the first born children, later born children were more likely to have lower gestational age (39 vs. 40 weeks), higher birth weight (3298 vs. 3259g), and were less likely to be overdue births (2.2% vs. 4.9%) (Table 3)

The median birth interval was 3.0 (mean 3.7, SD 2.4, mode 1.9) years, and the median interpregnancy-interval was 2.2 (mean 2.9, SD 2.4, mode 1.1) years. The median birth interval decreased as the number of siblings increased, from 3.1 years for clusters with two children to 2.8 for clusters with four children, to 2.0 for clusters with five or more children (Figure 3). Similarly, the median interpregnancy-interval decreased from 2.3 years for clusters with two children to 2.0 years for clusters with four children, to 1.3 years for clusters with five or more children (Supplementary Figure 1).

Figure 1: Flow diagram of participant inclusion in ECHILD-oCSib cohort
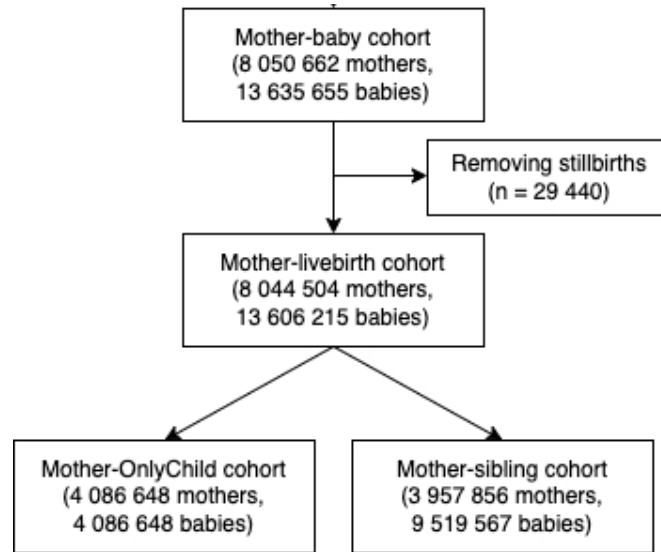
Discussion

We established a nationwide cohort of 4.1 million MoC clusters and 4.0 million MSib clusters, covering 87% of livebirths in England between April 1997 and January 2021. Through continuous linkage with longitudinal HES and NPD datasets, the cohort provides comprehensive insights into the health, education, and social care service use of each individual, spanning both mothers and the children they gave birth to. The unique combination of this extensive data resource and the well-defined cohort structure renders it valuable for addressing a diverse range of research questions.

First, the mother-baby structure provides a robust framework for investigating intergenerational effects of maternal exposures on children's outcomes measured in education, health or social care records [37, 42, 45, 46]. Extensive previous research has suggested the significant role of maternal factors in children's short- and long-term outcomes, spanning from birth through childhood to adolescent and adulthood [47–50]. For example, linked mother-baby HES has been used to examine the effects of a range of pre-pregnancy psychosocial risk factors on birth weight, injury admission and infant mortality [49]. Our data enable the examination of how the conditions and events experienced by mothers prior to and during pregnancy may shape the trajectories of all of their children, with special emphasis on health, education and social care use.

Second, this cohort enables comparison between MoC and MSib clusters, offering insights at both the child and mother levels. At the child level, the cohort facilitates research into whether having siblings is a risk/protective factor, or a potential effect modifier [6–8, 16, 51]. At the mother level, the cohort facilitates research into examining whether the number of children (or parity) a mother has is a potential risk or protective factor for outcomes, or if it serves as a potential effect modifier [52–56]. For example, for children, it has been suggested that having siblings promotes a more healthy weight status, while only-children are at a higher risk of overweight and obesity [6], and of adverse lipid profile [7]. Yu et al. found that sibling additions were associated with cognitive development, but only to first and second born children, but not later born children [51]. Despite the efforts put into this research field, uncertainty remains [51]. For instance, A Dutch study showed that the psychological wellbeing of mothers declines more rapidly than in childless women [53], whereas another study found an inverse association between having more children and risk of suboptimal mental health in White mothers aged 65 years or older [52]. A US study showed that multiparity is associated with poorer cardiovascular health [54], which was similar to a finding in UK Biobank [56], but in contrast to a null association between number of children and cardiovascular risk factors found in NHANES [55].

Third, the MSib cohort will be instrumental in facilitating examination of sibling dynamics and their impact on health, education, and social care involvement. The study of sibling characteristics is essential for gaining insights into the dynamics within family structures and their broader implications on individual development. On one hand, siblings contribute to establishment of social networks and emotional support; while on the other hand, they may dilute the care, support and resources provided by caregivers. A 30-year longitudinal study in the United States suggested the time-dependent nature of family resource-dilution processes; they revealed that having an older sibling is beneficial for sociobehavioral development, but gaining a younger sibling increases behavioural problems for some first-born children [51]. Investigating sibling relationships provides a unique lens through which researchers can comprehend the influence of familial environments on social, emotional, and cognitive development. By exploring siblings dynamics, researchers can inform the design of interventions and support systems that cater to the diverse needs arising from sibling relationships.

Fourth, sibling studies can be implemented within the MSib cohort, disentangling the influence of genetic and environmental confounders, thereby enhancing the potential for establishing causal inferences regarding exposure-outcome associations [57–59]. Conventional epidemiological study designs are prone to unmeasured genetic and environmental confounders, which may distort the association estimates.

Table 2: Characteristics of children in the mother-only-child (MoC) and mother-sibling (MSib) cohorts

	Children in MoC cohort	Children in MSib cohort						Total	Standardised mean difference (95%CI) between MoC and MSib children
		Sibship size = 2	Sibship size = 3	Sibship size = 4	Sibship size = 5	Sibship ize ≥ 6	Children with siblings		
	n = 4086648 (30.0%)	n = 5631314 (41.4%)	n = 2465997 (18.1%)	n = 906060 (6.6%)	n = 315990 (2.3%)	n = 200206 (1.5%)	n = 9519567 (70.0%)	n = 13606215 (100.0%)	
Male sex, n(%)	2079237 (50.9%)	2880700 (51.2%)	1271899 (51.6%)	462069 (51.0%)	159925 (50.6%)	101336 (50.7%)	4875929 (51.2%)	6955166 (51.1%)	0.007 (0.005, 0.008)
Ethnicities									0.187 (0.186, 0.188)
Asian	292432 (10.3%)	426679 (9.8%)	271574 (14.1%)	125163 (17.8%)	42750 (17.5%)	18483 (11.9%)	884649 (12.0%)	1177081 (11.5%)	
Black	162520 (5.7%)	186211 (4.3%)	121038 (6.3%)	51879 (7.4%)	19208 (7.8%)	11994 (7.7%)	390330 (5.3%)	552850 (5.4%)	
Mixed	152267 (5.3%)	193974 (4.5%)	81867 (4.3%)	29415 (4.2%)	10765 (4.4%)	6736 (4.3%)	322757 (4.4%)	475024 (4.6%)	
White	2240465 (78.7%)	3544697 (81.5%)	1445046 (75.3%)	496169 (70.6%)	172096 (70.3%)	117933 (76.0%)	5775941 (78.3%)	8016406 (78.4%)	
Unknown	1238964 (30.3%)	1279753 (22.7%)	546472 (22.2%)	203434 (22.5%)	71171 (22.5%)	45060 (22.5%)	2145890 (22.5%)	3384854 (24.9%)	
Maternal Index of multiple deprivation categories									0.105 (0.104, 0.107)
Quintile 1 (most deprived)	991941 (24.4%)	1263312 (22.5%)	788331 (32.1%)	386088 (42.7%)	156951 (49.8%)	106715 (53.4%)	2701397 (28.5%)	3693338 (27.3%)	
Quintile 2	882832 (21.8%)	1158159 (20.6%)	537191 (21.8%)	206330 (22.8%)	72123 (22.9%)	45002 (22.5%)	2018805 (21.3%)	2901637 (21.4%)	
Quintile 3	757712 (18.7%)	1068541 (19.0%)	418057 (17%)	132575 (14.7%)	40372 (12.8%)	23626 (11.8%)	1683171 (17.7%)	2440883 (18.0%)	
Quintile 4	676437 (16.7%)	1013771 (18.0%)	354100 (14.4%)	94000 (10.4%)	25316 (8.0%)	13753 (6.9%)	1500940 (15.8%)	2177377 (16.1%)	
Quintile 5 (least deprived)	748818 (18.5%)	1112681 (19.8%)	361738 (14.7%)	84757 (9.4%)	20428 (6.5%)	10564 (5.3%)	1590168 (16.7%)	2338986 (17.3%)	
Unknown	28908 (0.7%)	14850 (0.3%)	6580 (0.3%)	2310 (0.3%)	800 (0.3%)	546 (0.3%)	25806 (0.3%)	54714 (0.4%)	
Maternal age* (years), mean (SD)	30.02 (5.99)	29.65 (5.57)	28.36 (5.83)	27.41 (5.95)	27.12 (6.04)	27.62 (6.24)	28.98 (5.77)	29.28 (5.85)	0.178 (0.177, 0.179)
<20	158475 (4.6%)	195488 (4.0%)	143554 (6.6%)	72074 (9.0%)	28733 (10.3%)	17388 (9.8%)	457237 (5.5%)	615712 (5.2%)	0.186 (0.185, 0.188)
20-24	500232 (14.7%)	751482 (15.2%)	461451 (21.2%)	205399 (25.7%)	74846 (26.9%)	44068 (24.8%)	1537246 (18.4%)	2037478 (17.3%)	
25-29	880091 (25.8%)	1379567 (27.9%)	636632 (29.3%)	233900 (29.2%)	79536 (28.5%)	48962 (27.6%)	2378597 (28.4%)	3258688 (27.7%)	
30-34	1053499 (30.9%)	1625854 (32.9%)	586956 (27.0%)	181337 (22.7%)	59470 (21.3%)	39666 (22.3%)	2493283 (29.8%)	3546782 (30.1%)	
35-39	641648 (18.8%)	833673 (16.9%)	291742 (13.4%)	88494 (11.1%)	29336 (10.5%)	21753 (12.2%)	1264998 (15.1%)	1906646 (16.2%)	
40-44	168105 (4.9%)	147212 (3.0%)	52949 (2.4%)	17971 (2.2%)	6392 (2.3%)	5407 (3.0%)	229931 (2.7%)	398036 (3.4%)	
>=45	10516 (0.3%)	6844 (0.1%)	2396 (0.1%)	827 (0.1%)	365 (0.1%)	342 (0.2%)	10774 (0.1%)	21290 (0.2%)	
Unknown	674082 (16.5%)	691194 (12.3%)	290317 (11.8%)	106058 (11.7%)	37312 (11.8%)	22620 (11.3%)	1147501 (12.1%)	1821583 (13.4%)	

Continued

Table 2: Continued

	Children in MoC cohort	Children in MSib cohort						Total	Standardised mean difference (95%CI) between MoC and MSib children
		Sibship size = 2	Sibship size = 3	Sibship size = 4	Sibship size = 5	Sibship ize ≥ 6	Children with siblings		
	n = 4086648 (30.0%)	n = 5631314 (41.4%)	n = 2465997 (18.1%)	n = 906060 (6.6%)	n = 315990 (2.3%)	n = 200206 (1.5%)	n = 9519567 (70.0%)	n = 13606215 (100.0%)	
Gestational age at birth (weeks), median (interquartile range)	39 (38, 40)	39 (38, 40)	39 (38, 40)	39 (38, 40)	39 (38, 40)	39 (38, 40)	39 (38, 40)	39 (38, 40)	0.048 (0.047, 0.050)
<= 27	7478 (0.3%)	8377 (0.2%)	4372 (0.2%)	1909 (0.3%)	807 (0.3%)	557 (0.4%)	16022 (0.2%)	23500 (0.2%)	0.045 (0.044, 0.046)
28-31	20029 (0.7%)	25748 (0.6%)	13030 (0.7%)	5577 (0.8%)	2372 (1.0%)	1546 (1.0%)	48273 (0.7%)	68302 (0.7%)	
32-36	167628 (5.7%)	245424 (5.9%)	126643 (6.9%)	52952 (7.7%)	20648 (8.6%)	13714 (9%)	459381 (6.5%)	627009 (6.3%)	
37-41	2626033 (89.5%)	3750281 (89.9%)	1642561 (89.1%)	604590 (88.2%)	209556 (87.1%)	132190 (86.4%)	6339178 (89.4%)	8965211 (89.4%)	
>= 42	112909 (3.8%)	141071 (3.4%)	56669 (3.1%)	20368 (3.0%)	7184 (3.0%)	4980 (3.3%)	230272 (3.2%)	343181 (3.4%)	
Unknown	1152571 (28.2%)	1460413 (25.9%)	622722 (25.3%)	220664 (24.4%)	75423 (23.9%)	47219 (23.6%)	2426441 (25.5%)	3579012 (26.3%)	
Birth weight (g), mean (SD)	3252.05 (517.29)	3285.34 (513.01)	3250.85 (530.55)	3210.49 (538.2)	3184.06 (546.48)	3188.425 (547.67)	3263.6 (522.9)	3260.23 (521.3)	0.029 (0.028, 0.031)
<1500	31404 (1.0%)	35048 (0.8%)	17400 (0.9%)	7128 (1.0%)	2795 (1.1%)	1854 (1.1%)	64225 (0.8%)	95629 (0.9%)	0.051 (0.050, 0.053)
1500-1999	39101 (1.3%)	55683 (1.2%)	29354 (1.5%)	12563 (1.7%)	4912 (1.9%)	3114 (1.9%)	105626 (1.4%)	144727 (1.4%)	
2000-2499	145238 (4.6%)	201403 (4.5%)	106918 (5.4%)	45044 (6.1%)	17256 (6.7%)	10961 (6.7%)	381582 (5.0%)	526820 (4.9%)	
2500-2999	584429 (18.7%)	766564 (17.2%)	371834 (18.8%)	152510 (20.8%)	56333 (21.9%)	35107 (21.4%)	1382348 (18.2%)	1966777 (18.4%)	
3000-3499	1267376 (40.6%)	1761738 (39.5%)	765319 (38.7%)	283635 (38.7%)	99107 (38.5%)	62974 (38.4%)	2972773 (39.2%)	4240149 (39.6%)	
3500-3999	952924 (30.5%)	1460925 (32.7%)	607548 (30.8%)	207660 (28.3%)	69084 (26.8%)	44974 (27.4%)	2390191 (31.5%)	3343115 (31.2%)	
4000-4499	94125 (3.0%)	162804 (3.6%)	69041 (3.5%)	22698 (3.1%)	7247 (2.8%)	4643 (2.8%)	266433 (3.5%)	360558 (3.4%)	
4500-4999	8229 (0.3%)	15714 (0.4%)	7023 (0.4%)	2181 (0.3%)	769 (0.3%)	419 (0.3%)	26106 (0.3%)	34335 (0.3%)	
>= 5000	1503 (0.1%)	2158 (0.1%)	965 (0.1%)	301 (0.1%)	97 (0.1%)	51 (0.1%)	3572 (0.1%)	5075 (0.1%)	
Unknown	962319 (23.5%)	1169277 (20.8%)	490595 (19.9%)	172340 (19.0%)	58390 (18.5%)	36109 (18.0%)	1926711 (20.2%)	2889030 (21.2%)	
Neonatal care admission									0.056 (0.055, 0.057)
Normal care	2342795 (86.2%)	3331307 (87.6%)	1488677 (87.7%)	554409 (87.5%)	192616 (86.9%)	123811 (86.9%)	5690820 (87.5%)	8033615 (87.1%)	
Special care	304313 (11.2%)	379784 (10.1%)	164133 (9.7%)	61142 (9.7%)	22230 (10.0%)	14145 (9.9%)	641434 (9.9%)	945747 (10.3%)	
L1 intensive care	43541 (1.6%)	58411 (1.5%)	28114 (1.7%)	11373 (1.8%)	4410 (2.0%)	2921 (2.0%)	105229 (1.6%)	148770 (1.6%)	
L2 intensive care	28775 (1.1%)	35327 (0.9%)	16604 (1.1%)	6605 (1.0%)	2449 (1.1%)	1666 (1.2%)	62651 (1.0%)	91426 (1.0%)	
Unknown	1367224 (33.5%)	1826485 (32.4%)	768469 (31.2%)	272531 (30.1%)	94285 (29.8%)	57663 (28.8%)	3019433 (31.7%)	4386657 (32.2%)	

When calculating percentage for Unknown category, the denominator was all total number of babies; when calculating percentage for other categories, the denominator was the total number of babies excluding the unknown category. *: the same mother is included multiple times in the maternal characteristics descriptive.

Table 3: Characteristics of children in the mother-sibling (Msib) cohorts by birth order (first born vs. later born) (removing the records where the first delivery was a multiple delivery)

	First born children n = 3898717 (41.7%)	Later born children n = 5457990 (58.3%)	Total n = 9356707 (100.0%)	Standardised mean difference (95%CI) between first born and later born
Male sex, n (%)	2000284 (51.3%)	2795032 (51.2%)	4795316 (51.3%)	0.002 (0.001, 0.003)
Ethnicities				0.277 (0.276, 0.278)
Asian	309237 (11.2%)	564952 (12.6%)	874189 (12%)	
Black	139507 (5.0%)	243737 (5.4%)	383244 (5.3%)	
Mixed	112998 (4.1%)	204251 (4.5%)	317249 (4.4%)	
White	2201216 (79.7%)	3479318 (77.5%)	5680534 (78.3%)	
Unknown	1135759 (29.1%)	965732 (17.7%)	2101491 (22.5%)	
Index of multiple deprivation categories				0.081 (0.080, 0.083)
Quintile 1 (most deprived)	1076228 (27.7%)	1590426 (29.2%)	2666654 (28.6%)	
Quintile 2	947472 (24.4%)	1314712 (24.1%)	2262184 (24.2%)	
Quintile 3	852241 (21.9%)	1133733 (20.8%)	1985974 (21.3%)	
Quintile 4	777559 (20.0%)	1015838 (18.6%)	1793397 (19.2%)	
Quintile 5 (least deprived)	716799 (18.4%)	935514 (17.2%)	1652313 (17.7%)	
Unknown	12456 (0.3%)	11070 (0.2%)	23526 (0.3%)	
Maternal age (years), mean (SD)	26.6 (5.5)	30.5 (5.4)	28.9 (5.8)	0.713 (0.711, 0.714)
<20	388890 (11.6%)	62925 (1.3%)	451815 (5.5%)	0.703 (0.701, 0.704)
20-24	840523 (25.1%)	678016 (13.9%)	1518539 (18.5%)	
25-29	1026440 (30.6%)	1314876 (27.0%)	2341316 (28.5%)	
30-34	835713 (24.9%)	1604924 (33.0%)	2440637 (29.7%)	
35-39	241913 (7.2%)	990879 (20.4%)	1232792 (15.0%)	
40-44	18333 (0.5%)	204016 (4.2%)	222349 (2.7%)	
>=45	503 (0.1%)	8623 (0.2%)	9126 (0.1%)	
Unknown	546402 (14.0%)	593731 (10.9%)	1140133 (12.2%)	
Gestational age at birth (weeks), median (interquartile range)	40 (39, 41)	39 (38, 40)	39 (38, 40)	0.212 (0.210, 0.213)
<=27	5190 (0.2%)	8355 (0.2%)	13545 (0.2%)	0.155 (0.153, 0.156)
28-31	14777 (0.5%)	24921 (0.6%)	39698 (0.6%)	
32-36	145994 (5.1%)	250682 (6.2%)	396676 (5.7%)	
37-41	2569704 (89.3%)	3697705 (90.9%)	6267409 (90.2%)	
>=42	141428 (4.9%)	88267 (2.2%)	229695 (3.3%)	
Unknown	1021624 (26.2%)	1388060 (25.4%)	2409684 (25.8%)	
Birth weight (g), mean(SD)	3259 (495)	3298 (512)	3282 (505)	0.113 (0.112, 0.114)
<1500	21571 (0.7%)	29853 (0.7%)	51424 (0.7%)	0.108 (0.107, 0.110)
1500-1999	31831 (1.0%)	51767 (1.2%)	83598 (1.1%)	
2000-2499	136605 (4.4%)	194755 (4.5%)	331360 (4.5%)	
2500-2999	592619 (19.2%)	739753 (17.0%)	1332372 (17.9%)	
3000-3499	1273434 (41.3%)	1682794 (38.7%)	2956228 (39.7%)	
3500-3999	927067 (30.0%)	1459805 (33.5%)	2386872 (32.1%)	
4000-4499	93433 (3.0%)	172676 (4.0%)	266109 (3.6%)	
4500-4999	8149 (0.3%)	17940 (0.4%)	26089 (0.4%)	
>=5000	1172 (0.1%)	2156 (0.1%)	3328 (0.1%)	
Unknown	812836 (20.8%)	1106491 (20.3%)	1919327 (20.5%)	
Neonatal care admission				0.065 (0.064, 0.067)
Normal care	2496099 (88.5%)	3126876 (87.8%)	5622975 (88.1%)	
Special care	263689 (9.3%)	344525 (9.7%)	608214 (9.5%)	
L1 intensive care	37949 (1.3%)	56407 (1.6%)	94356 (1.5%)	
L2 intensive care	23162 (0.8%)	33345 (0.9%)	56507 (0.9%)	
Unknown	1077818 (27.6%)	1896837 (34.8%)	2974655 (31.8%)	

When calculating percentage for Unknown category, the denominator was all total number of babies; when calculating percentage for other categories, the denominator was the total number of babies excluding the unknown category. *: the same mother is included multiple times in the maternal characteristics descriptive. In this table, we removed the records where the first delivery was a multiple delivery.

Figure 2: The percentage of multiple births by the size of mother-sibling cluster
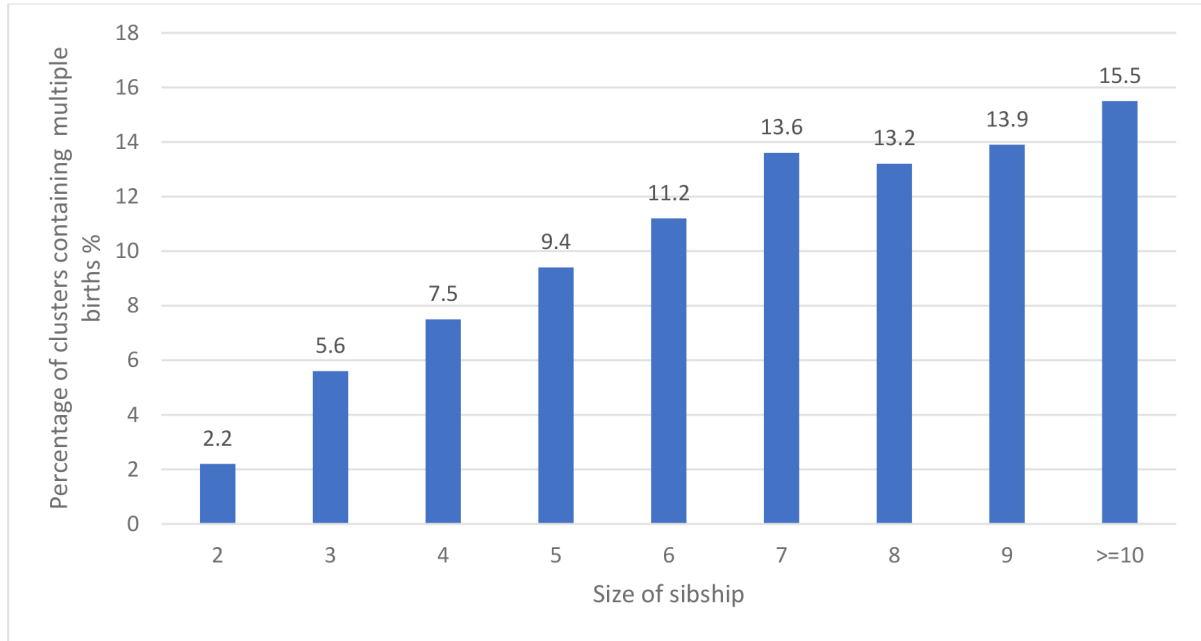
The number of siblings is the number of siblings in each mother-sibling cluster by January 2022. The numerator is the number of mother-sibling clusters that have ever had multiple births between April 1997 and January 2022. The denominator is the number of the mother-sibling clusters that have the according number of siblings in January 2022. We included only livebirths when counting the number of siblings.

Figure 3: Birth intervals in mother-siblings cohort by the number of siblings
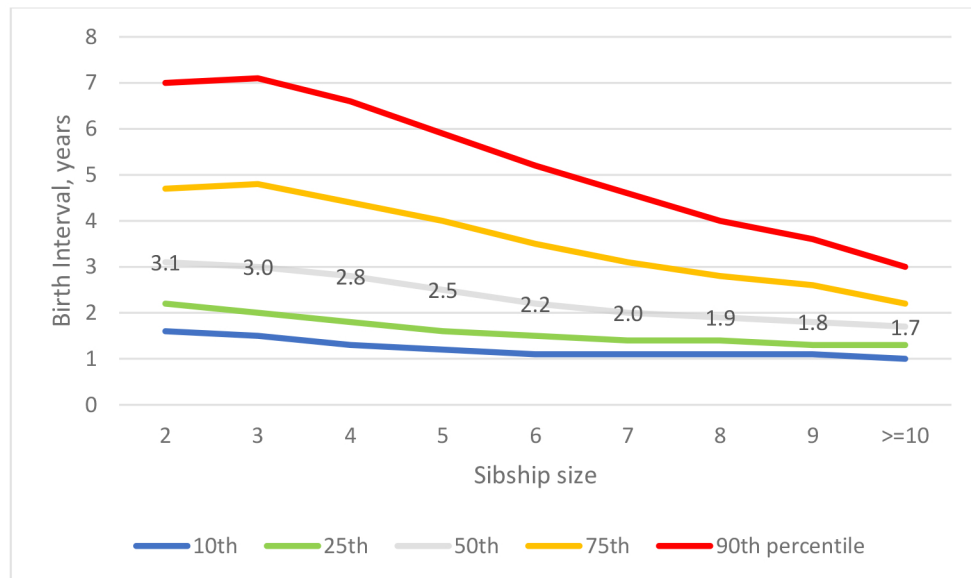
Birth interval is the time period between one delivery and the next delivery. We excluded the first delivery (regardless of whether it was a singleton birth or multiple birth) when calculating the summary statistics of birth interval, as they were always zero by definition. The figure summarised the values for individual children, not by sibling clusters. The numbers in the plots represent the median values (50th percentile).

In contrast, sibling comparison designs leverage the shared genetic background, lifestyle and environmental factors among siblings, serving to minimise such bias by comparing siblings that are discordant for particular exposures or outcomes [60]. It is acknowledged that new sources of bias, including birth order, age differences, gestational age, and maternal and paternal age at birth, may still be present, necessitating additional adjustments in the analytical approach.

Fifth, the structure of this cohort allows research into interactions between maternal factors and sibling dynamics, thus yielding a more comprehensive picture of how family-level factors influence child outcomes. This holistic approach not only contributes to a more comprehensive understanding of the complexities involved within families but also affords greater precision in drawing meaningful conclusions regarding the multifaceted relationships within family structures and their impact on health, education, and overall well-being. Currently, research evidence in this area is lacking.

Several potential sources of bias warrant careful consideration. Firstly, the siblings identified in our dataset are based on children born to the same mother, and we are not able to distinguish whether they share the same father (full sibling) or have different fathers (half-sibling). Secondly, the identified siblings may not necessarily be the individuals with whom a specific child grew up, as circumstances such as adoption or stepsibling relationships could alter family composition; for similar reasons, the linked mothers may not necessarily be the caregivers. However, these factors are anticipated to be validated in the future through linkage to Unique Property Reference Numbers (UPRN) [61]. Thirdly, the inclusion of an external, independent dataset to corroborate sibling status is recommended for robust validation. Fourthly, HES only captures birth records in NHS hospitals in England between April 1997 and January 2022, therefore, we may miss births before April 1997, or after January 2022, or outside NHS hospitals, or outside England (for example, if a mother immigrated to England with her existing child(ren) born in other countries), which introduces bias in underestimating the number of children a mother actually had. The size of ECHILD-MoC may be overestimated, while the size of ECHILD-MSib may be underestimated as well as the size of sibship. A previous UK study demonstrated that mothers who had out of NHS hospital deliveries were more likely to be older, white, more affluent and multiparous [62], which might imply possibility of selection bias. Fifthly, as the cohort is linked to national datasets with constant updates of data, we expect some mothers to give birth to more babies in the future, thus meaning that some individuals in the MoC cohort may be moved to the MSib cohort, and sizes of sibships in the MSib cohort will increase over time. Sixthly, NHS England apply opt outs for overall 5.4% of English population (1.3% of children aged 0-9 years and 8.5% for those aged 30-39 years). Seventhly, there is missing data in our cohort, as is in other administrative databases. Some variables have higher proportions of missing data than others, such as ethnicity, gestational age, birth weight, neonatal care admission. Data completeness has been significantly increased via data imputation based on mother-baby pairs and sibling status (for some delivery/birth related variables) and mapping across multiple medical records (for ethnicity), however, missingness is still high for some variables. These limitations may contribute to selection and information biases and should be taken into account when interpreting findings based on these data. Vigilance in addressing these potential biases will ensure a more accurate and nuanced understanding of the complex familial dynamics under investigation.

Data access

ECHILD-MoC and ECHILD-MSib data is available to accredited researchers, as part of ECHILD database. ECHILD is accessible via ONS Secure Research Service, a trusted research environment. Researchers wishing to access the ECHILD Database must apply to the ECHILD Data Access Committee. For more information on application process, please contact ich.echild@ucl.ac.uk.

Statements

Acknowledgement

Ethics statement

Ethical approval for the ECHILD project was granted by the National Research Ethics Service (17/LO/1494), NHS Health Research Authority Research Ethics Committee (20/EE/0180 and 21/SW/0159) and is overseen by the UCL Great Ormond Street Institute of Child Health's Joint Research and Development Office (20PE16).

Conflict of interests statement

The authors report no conflicts of interest.

Publication consent

The authors have gained publication consent to publish and openly share the data to accredited researchers via a safe research environment.

Funding statement

Data availability statement

ECHILD-MoC and ECHILD-MSib data is available to accredited researchers, as part of ECHILD database. ECHILD is accessible via ONS Secure Research Service, a trusted research environment. Researchers wishing to access the ECHILD Database must apply to the ECHILD Data Access Committee. For more information on application process, please contact ich.echild@ucl.ac.uk.

References

1. statista. Number of families iin the United Kingdom from 1996 to 2022, by number of dependent children [Internet]. Soc. Demogr. 2023 May [cited 2023 16]; Available from: https://www.statista.com/statistics/734771/family-sizes-uk/.

2. Gilligan M, Stocker CM, Jewsbury Conger K. Sibling Relationships in Adulthood: Research Findings and New Frontiers. J. Fam. Theory Rev. 2020 Sep.;12:305–20. https://doi.org/10.1111/jftr.12385

3. Keenan K, Barclay K, Goisis A. Health outcomes of only children across the life course: An investigation using Swedish register data. Popul. Stud. 2023 Jan.;77:71–90. https://doi.org/10.1080/00324728.2021.2020886

4. Kracht CL, Sisson SB. Sibling influence on children's objectively measured physical activity: a meta-analysis and systematic review. BMJ Open Sport Exerc. Med. 2018 Jul.;4:e000405. https://doi.org/10.1136/bmjsem-2018-000405

5. Tian X, Von Cramon-Taubadel S. Are only children in China more likely to be obese/overweight than their counterparts with siblings? Econ. Hum. Biol. 2020 May;37:100847. https://doi.org/10.1016/j.ehb.2020.100847

6. Bohn C, Vogel M, Poulain T, Hiemisch A, Kiess W, Körner A. Having siblings promotes a more healthy weight status—Whereas only children are at greater risk for higher BMI in later childhood. PLOS ONE 2022 Jul.;17:e0271676. https://doi.org/10.1371/journal.pone.0271676

7. Cai L, Ma B, Lin L, Chen Y, Yang W, Ma J, et al. The differences of lipid profiles between only children and children with siblings: A national survey in China. Sci. Rep. 2019 Feb.;9:1441. https://doi.org/10.1038/s41598-018-37695-0

8. Kwan Y, Ip W. Life Satisfaction, Perceived Health, Violent and Altruistic Behaviour of Hong Kong Chinese Adolescents: Only Children Versus Children with Siblings. Child Indic. Res. 2009 Dec.;2:375–89. https://doi.org/10.1007/s12187-009-9041-y

9. Chen Z, Liu RX. Comparing Adolescent Only Children with Those Who Have Siblings on Academic Related Outcomes and Psychosocial Adjustment. Child Dev. Res. 2014 Jan.;2014:1–10. https://doi.org/10.1155/2014/578289

10. Howe N, Recchia H. Sibling Relationships as a Context for Learning and Development. Early Educ. Dev. 2014 Feb.;25:155–9. https://doi.org/10.1080/10409289.2014.857562

11. McHale SM, Updegraff KA, Whiteman SD. Sibling Relationships and Influences in Childhood and Adolescence. J. Marriage Fam. 2012 Oct.;74:913–30. https://doi.org/10.1111/j.1741-3737.2012.01011.x

12. Fukuya Y, Fujiwara T, Isumi A, Doi S, Ochi M. Association of Birth Order With Mental Health Problems, Self-Esteem, Resilience, and Happiness Among Children: Results From A-CHILD Study. Front. Psychiatry 2021 Apr.;12:638088. https://doi.org/10.3389/fpsyt.2021.638088

13. Carballo JJ, García-Nieto R, Álvarez-García R, Caro-Cañizares I, López-Castromán J, Muñoz-Lorenzo L, et al. Sibship size, birth order, family structure and childhood mental disorders. Soc. Psychiatry Psychiatr. Epidemiol. 2013 Aug.;48:1327–33. https://doi.org/10.1007/s00127-013-0661-7

14. Dhamrait G, Fletcher T, Foo D, Taylor CL, Pereira G. The effects of birth spacing on early childhood development in high-income nations: A systematic review. Front. Pediatr. 2022 Nov.;10:851700. https://doi.org/10.3389/fped.2022.851700

15. Bliznashka L, Jeong J. Investigating the direct and indirect associations between birth intervals and child growth and development: A cross-sectional analysis of 13 Demographic and Health Surveys. SSM - Popul. Health 2022 Sep.;19:101168. https://doi.org/10.1016/j.ssmph.2022.101168

16. Lawson DW, Mace R. Siblings and childhood mental health: Evidence for a later-born advantage. Soc. Sci. Med. 2010 Jun.;70:2061–9. https://doi.org/10.1016/j.socscimed.2010.03.009

17. French R, Sariaslan A, Larsson H, Kneale D, Leckie G. Estimating the Importance of Families in Modeling Educational Achievement Using Linked Swedish Administrative Data. J. Res. Educ. Eff. 2023 Jan.;16:106–33. https://doi.org/10.1080/19345747.2022.2054480

18. Choi S, Taiji R, Chen M, Monden C. Cohort Trends in the Association Between Sibship Size and Educational

Attainment in 26 Low-Fertility Countries. Demography 2020 Jun.;57:1035–62. https://doi.org/10.1007/s13524-020-00885-5

19. Quintana Mariñez MG, Chakkera M, Ravi N, Ramaraju R, Vats A, Nair AR, et al. The Other Sibling: Mental Health Effects on a Healthy Sibling of a Child With a Chronic Disease: A Systematic Review. Cureus [Internet] 2022 Sep. [cited 2023 19]; Available from: https://www.cureus.com/articles/110965-the-other-sibling-mental-health-effects-on-a-healthy-sibling-of-a-child-with-a-chronic-disease-a-systematic-review. https://doi.org/10.7759/cureus.29042

20. Hanvey I, Malovic A, Ntontis E. Glass children: The lived experiences of siblings of people with a disability or chronic illness. J. Community Appl. Soc. Psychol. 2022 Sep.;32:936–48. https://doi.org/10.1002/casp.2602

21. Fleary SA, Heffer RW. Impact of Growing Up with a Chronically Ill Sibling on Well Siblings' Late Adolescent Functioning. ISRN Fam. Med. 2013 Jan.;2013:1–8. https://doi.org/10.5402/2013/737356

22. Reed T, Carmelli D, Christian JC, Selby JV, Fabsitz RR. The NHLBI male veteran twin study data. Genet. Epidemiol. 1993 Jan.;10:513–7. https://doi.org/10.1002/gepi.1370100630

23. Lichtenstein P, De Faire U, Floderus B, Svartengren M, Svedberg P, Pedersen NL. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. J. Intern. Med. 2002 Sep.;252:184–205. https://doi.org/10.1046/j.1365-2796.2002.01032.x

24. Skytthe A, Ohm Kyvik K, Vilstrup Holm N, Christensen K. The Danish Twin Registry. Scand. J. Public Health 2011 Jul.;39:75–8. https://doi.org/10.1177/1403494810387966

25. Nilsen TS, Knudsen GP, Gervin K, Brandt I, Røysamb E, Tambs K, et al. The Norwegian Twin Registry from a Public Health Perspective: A Research Update. Twin Res. Hum. Genet. 2013 Feb.;16:285–95. https://doi.org/10.1017/thg.2012.117

26. Li L, Gao W, Yu C, Lv J, Cao W, Zhan S, et al. The Chinese National Twin Registry: an update. Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud. 2013 Feb.;16:86–90. https://doi.org/10.1017/thg.2012.148

27. Silventoinen K, Jelenkovic A, Sund R, Honda C, Aaltonen S, Yokoyama Y, et al. The CODATwins Project: The Cohort Description of Collaborative Project of Development of Anthropometrical Measures in Twins to Study Macro-Environmental Variation in Genetic and Environmental Effects on Anthropometric Traits. Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud. 2015 Aug.;18:348–60. https://doi.org/10.1017/thg.2015.29

28. Office of National Statistics. Birth Characteristics Dataset [Internet]. 2023 Jan. [cited 2024 23]; Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthcharacteristicsinenglandandwales

29. Koeppen-Schomerus G, Spinath FM, Plomin R. Twins and non-twin siblings: different estimates of shared environmental influence in early childhood. Twin Res. Off. J. Int. Soc. Twin Stud. 2003 Apr.;6:97–105. https://doi.org/10.1375/136905203321536227

30. Mönkediek B, Schulz W, Eichhorn H, Diewald M. Is there something special about twin families? A comparison of parenting styles in twin and non-twin families. Soc. Sci. Res. 2020 Aug.;90:102441. https://doi.org/10.1016/j.ssresearch.2020.102441

31. Bohm HV, Stewart MG, Healy AM. On the Autistic Spectrum Disorder concordance rates of twins and non-twin siblings. Med. Hypotheses 2013 Nov.;81:789–91. https://doi.org/10.1016/j.mehy.2013.08.019

32. Chen L, Cnattingius S, Nyman Iliadou A, Oberg AS. Cancer risks in twins and singletons from twin and non-twin families. Int. J. Cancer 2016 Mar.;138:1102–10. https://doi.org/10.1002/ijc.29866

33. Boone-Heinonen J, Biel FM, Marshall NE, Snowden JM. Maternal prepregnancy BMI and size at birth: race/ethnicity-stratified, within-family associations in over 500,000 siblings. Ann. Epidemiol. 2020 Jun.;46:49-56.e5. https://doi.org/10.1016/j.annepidem.2020.04.009

34. Class QA, Rickert ME, Larsson H, Lichtenstein P, D'Onofrio BM. Fetal growth and psychiatric and socioeconomic problems: population-based sibling comparison. Br. J. Psychiatry 2014 Nov.;205:355–61. https://doi.org/10.1192/bjp.bp.113.143693

35. Daley D, Jacobsen RH, Lange A-M, Sørensen A, Walldorf J. The economic burden of adult attention deficit hyperactivity disorder: A sibling comparison cost analysis. Eur. Psychiatry 2019 Sep.;61:41–8. https://doi.org/10.1016/j.eurpsy.2019.06.011

36. Chanfreau J, Barclay K, Keenan K, Goisis A. Sibling group size and BMI over the life course: Evidence from four British cohort studies. Adv. Life Course Res. 2022 Sep.;53:100493. https://doi.org/10.1016/j.alcr.2022.100493

37. Mc Grath-Lone L, Libuy N, Harron K, Jay MA, Wijlaars L, Etoori D, et al. Data Resource Profile: The Education and Child Health Insights from Linked Data (ECHILD) Database. Int. J. Epidemiol. 2022 Feb.;51:17–17f. https://doi.org/10.1093/ije/dyab149

38. Feng Q, Ireland G, Gilbert R, Harron K. Data Resource Profile: a national linked mother-baby cohort of health, education and social care data in England (ECHILD-MB) [In press]. Int J Epidemiol

39. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES

APC). Int. J. Epidemiol. 2017 Aug.;46:1093–1093i. https://doi.org/10.1093/ije/dyx015

40. Jay MA, Mc Grath-Lone L, Gilbert R. Data Resource: the National Pupil Database (NPD). Int. J. Popul. Data Sci. 2019 Mar.;4:1101. https://doi.org/10.23889/ijpds.v4i1.1101

41. Libuy N, Harron K, Gilbert R, Caulton R, Cameron E, Blackburn R. Linking education and hospital data in England: linkage process and quality. Int. J. Popul. Data Sci. [Internet] 2021 Sep. [cited 2023 20];6. Available from: https://ijpds.org/article/view/1671. https://doi.org/10.23889/ijpds.v6i1.1671

42. UCL ECHILD Group. ECHILD User Guide. ECHILD User Guide 2023 Jul.

43. Harron K, Gilbert R, Cromwell D, Van Der Meulen J. Linking Data for Mothers and Babies in De-Identified Electronic Health Data. PLOS ONE 2016 Oct.;11:e0164667. https://doi.org/10.1371/journal.pone.0164667

44. Ireland G, Jay M, Feng Q, Harron K, Grant C, Wijlaars L, et al. Linkage of administrative family court care proceedings and hospital records for mothers in England: linkage accuracy and cumulative incidence of family court care proceedings after a first live birth [In press]. Int J Pop Data Sci 2024.

45. Zylbersztejn A, Lewis K, Nguyen V, Matthews J, Winterburn I, Karwatowska L, et al. Evaluation of variation in special educational needs provision and its impact on health and education using administrative records for England: umbrella protocol for a mixed-methods research programme. BMJ Open 2023 Nov.;13:e072531. https://doi.org/10.1136/bmjopen-2023-072531

46. John A, Friedmann Y, DelPozo-Banos M, Frizzati A, Ford T, Thapar A. Association of school absence and exclusion with recorded neurodevelopmental disorders, mental disorders, or self-harm: a nationwide, retrospective, electronic cohort study of children and young people in Wales, UK. Lancet Psychiatry 2022 Jan.;9:23–34. https://doi.org/10.1016/S2215-0366(21)00367-9

47. Likhar A, Patil MS. Importance of Maternal Nutrition in the First 1,000 Days of Life and Its Effects on Child Development: A Narrative Review. Cureus 2022 Oct.;14:e30083. https://doi.org/10.7759/cureus.30083

48. Hardie JH, Landale NS. Profiles of Risk: Maternal Health, Socioeconomic Status, and Child Health. J. Marriage Fam. 2013 Jun.;75:651–66. https://doi.org/10.1111/jomf.12021

49. Harron K, Gilbert R, Fagg J, Guttmann A, Van Der Meulen J. Associations between pre-pregnancy psychosocial risk factors and infant outcomes: a population-based cohort study in England. Lancet Public Health 2021 Feb.;6:e97–105. https://doi.org/10.1016/S2468-2667(20)30210-3

50. Moog NK, Cummings PD, Jackson KL, Aschner JL, Barrett ES, Bastain TM, et al. Intergenerational transmission of the effects of maternal exposure to childhood maltreatment in the USA: a retrospective cohort study. Lancet Public Health 2023 Mar.;8:e226–37. https://doi.org/10.1016/S2468-2667(23)00025-7

51. Yu W, Yan HX. Effects of Siblings on Cognitive and Sociobehavioral Development: Ongoing Debates and New Theoretical Insights. Am. Sociol. Rev. 2023 Dec.; 88: 1002–30. https://doi.org/10.1177/00031224231210258

52. Van Den Broek T. Is having more children beneficial for mothers' mental health in later life? Causal evidence from the national health and aging trends study. Aging Ment. Health 2021 Oct.;25: 1950–8. https://doi.org/10.1080/13607863.2020.1774739

53. Kuipers YJ, Beeck EV, Cijsouw A, Van Gils Y. The impact of motherhood on the course of women's psychological wellbeing. J. Affect. Disord. Rep. 2021 Dec.;6:100216. https://doi.org/10.1016/j.jadr.2021.100216

54. Ogunmoroti O, Osibogun O, Kolade OB, Ying W, Sharma G, Vaidya D, et al. Multiparity is associated with poorer cardiovascular health among women from the Multi-Ethnic Study of Atherosclerosis. Am. J. Obstet. Gynecol. 2019 Dec.;221:631.e1-631.e16. https://doi.org/10.1016/j.ajog.2019.07.001

55. Neshteruk CD, Norman K, Armstrong SC, Cholera R, D'Agostino E, Skinner AC. Association between parenthood and cardiovascular disease risk: Analysis from NHANES 2011–2016. Prev. Med. Rep. 2022 Jun.;27:101820. https://doi.org/10.1016/j.pmedr.2022.101820

56. Magnus MC, Iliodromiti S, Lawlor DA, Catov JM, Nelson SM, Fraser A. Number of Offspring and Cardiovascular Disease Risk in Men and Women: The Role of Shared Lifestyle Characteristics. Epidemiology 2017 Nov.;28:880–8. https://doi.org/10.1097/EDE.0000000000000712

57. Petersen AH, Lange T. What Is the Causal Interpretation of Sibling Comparison Designs? Epidemiology 2020 Jan.;31:75–81. https://doi.org/10.1097/EDE.0000000000001108

58. Keyes KM, Smith GD, Susser E. On Sibling Designs: Epidemiology 2013 May;24:473–4. https://doi.org/10.1097/EDE.0b013e31828c7381

59. Frisell T, Öberg S, Kuja-Halkola R, Sjölander A. Sibling comparison designs: bias from non-shared confounders and measurement error. Epidemiol. Camb. Mass 2012 Sep.;23:713–20. https://doi.org/10.1097/EDE.0b013e31825fa230

60. Donovan SJ, Susser E. Commentary: Advent of sibling designs. Int. J. Epidemiol. 2011 Apr.;40:345–9. https://doi.org/10.1093/ije/dyr057

61. Johnson RD, Griffiths LJ, Hollinghurst JP, Akbari A, Lee A, Thompson DA, et al. Deriving household composition using population-scale electronic health record data-A reproducible methodology. PloS One 2021;16:e0248195. https://doi.org/10.1371/journal.pone.0248195

62. Birthplace in England Collaborative Group, Brocklehurst P, Hardy P, Hollowell J, Linsell L, Macfarlane A, et al. Perinatal and maternal outcomes by planned place of birth for healthy women with low risk pregnancies: the Birthplace in England national prospective cohort study. BMJ 2011 Nov.;343:d7400. https://doi.org/10.1136/bmj.d7400

Supplementary Figure 1: Inter-pregnancy interval in mother-siblings cohort by the sibship size
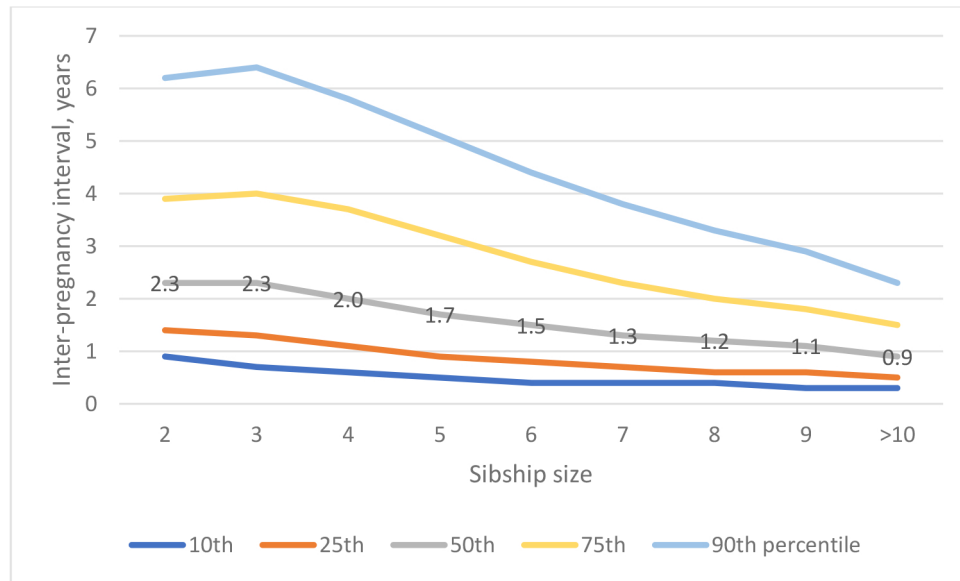
Inter-pregnancy interval is the time period between one delivery and the estimated conception date of next pregnancy. Conception date was estimated based on delivery date and gestation age; we used the average gestational age (39 weeks) if it was missing. We excluded the first delivery (regardless of whether it was a singleton birth or multiple birth) when calculating the summary statistics of inter-pregnancy interval, as they were always zero by definition. The figure summarised the values for individual children, not by sibling clusters. The numbers in the plots represent the median values (p50).

Supplementary Table 1: The codelists used to define stillbirth in Hospital Episode Statistics

Data source	Fields	Codelist
Diagnosis codes	ICD10	P95, Z371, Z373, Z374, Z376, Z377
HES specific fields	Discharge method	5
	Birth status	2-4

Supplementary Table 2: The number of boys in the mother-siblings cohort

	Number of boys in sibling cluster								
Sibship size	0	1	2	3	4	5	6	>=7	Total
2	651248 (23.1%)	1446101 (51.4%)	718308 (25.5%)	–	–	–	–	–	2815657 (100%)
3	99068 (12.1%)	291315 (35.4%)	313113 (38.1%)	118503 (14.4%)	–	–	–	–	821999 (100%))
4	14640 (6.5%)	54677 (24.1%)	81170 (35.8%)	58580 (25.9%)	17448 (7.7%)	–	–	–	226515 (100%))
5	2234 (3.5%)	10193 (16.1%)	18594 (29.4%)	18669 (29.5%)	10802 (17.1%)	2706 (4.3%)	–	–	63198 (100%)
6	373 (1.9%)	2051 (10.3%)	4526 (22.7%)	5743 (28.8%)	4528 (22.7%)	2261 (11.3%)	446 (2.2%)	–	19928 (100%)
7	57 (0.9%)	389 (6.0%)	1051 (16.1%)	1647 (25.3%)	1696 (26.0%)	1154 (17.7%)	438 (6.7%)	86 (1.3%)	6518 (100%)
8	13 (0.5%)	78 (3.2%)	253 (10.4%)	511 (20.9%)	607 (24.9%)	583 (23.9%)	288 (11.8%)	108 (4.4%)	2441 (100%)
>=9	3 (0.2%)	24 (1.5%)	74 (4.6%)	215 (13.4%)	313 (19.6%)	397 (24.8%)	308 (19.3%)	98 (10.4%)	1600 (100%)
Total	767636 (19.4%)	1804828 (45.6%)	1137089 (28.7%)	203868 (5.2%)	35394 (0.9%)	7101 (0.2%)	1480 (0.1%)	348 (0.1%)	3957856 (100%)

The number of siblings and the number of boys in each mother-sibling cluster were counted up to January 2022.