International Journal of Population Data Science

Journal Website: www.ijpds.org

# 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

Georgina Ireland[1,*], Linda Wijlaars[1], Matthew Jay[1], Qi Feng[1], Katie Harron[1], Claire Grant[2], and Ruth Gilbert[1]

Submission History

| | |
|---|---|
| Submitted: | 08/03/2024 |
| Accepted: | 06/04/2024 |
| Published: | 17/06/2024 |

[1]Institute of Child Health, UCL, 30 Guilford St, London WC1N 1EH
[2]Institute of Epidemiology and Health Care, UCL, 1-19 Torrington Place, London WC1E 7HB

## Abstract

### Introduction
Linkage of public law family court care proceedings (CP) data to all women giving birth in NHS hospitals in England allows calculation of the cumulative incidence of CP involvement for mothers with first children born.

### Objectives
To assess linkage accuracy and determine the 10-year cumulative incidence of CP after a first live birth (FLB) for population subgroups.

### Method
NHS England linked records for mothers in Cafcass (Children and Family Court Advisory and Support Service) involved in CP (2007-2021) to all mothers with a delivery in England using Hospital Episode Statistics (HES: 1997-21). We calculated match rates and assessed indirect evidence of potential false positive and missed links. We used survival analyses to estimate cumulative incidence of CP within 10 years overall and for five-year maternal age groups at first live birth.

### Results
Of 120,937 mothers involved in CP, 6.6% (n = 8,010) were excluded due to missing postcode or date of birth, or age <15 or >50. Of the remaining 112,927 mothers, 92,891 (82.8%) were linked to a HES delivery record. Match rates were lowest for mothers with an ethnic minority background, older at first case, or residing in Greater London, but improved over time.

Of 3,572,737 mothers with a FLB, 38,462 had CP involvement. The cumulative incidence of CP at 10 years from FLB was 1.31% (95% Confidence Interval [CI]; 1.29–1.32) overall and highest in mothers aged 15–19 years (6.79%, 95% CI: 6.69–6.89) and those living in the most deprived areas (2.47%, 95% CI: 2.43–2.51).

### Conclusion
One in 77 of all mothers and one in 15 aged less than 20 at first live birth were involved in CP within 10 years. Linkage error may underestimate the incidence of CP for mothers in London or with an ethnic minority background.

### Key points

- Overall, 82.8% of women recorded as a mother in Cafcass care proceedings were linked to a hospital delivery record.
- Match rates were lowest for mothers with an ethnic minority background, older age at first child, or residing in Greater London.
- 1.3% of all mothers (1 in 77) with a first birth were involved in care proceedings within 10 years and 6.8% (1 in 15) of mothers aged <20 at first live birth.

### Keywords
family court; care proceedings; data linkage; administrative data; health data

*Corresponding Author:
*Email Address:* g.ireland@ucl.ac.uk (Georgina Ireland)

https://doi.org/10.23889/ijpds.v9i2.2404

# Introduction

Involvement in public law family court care proceedings reflects serious breakdown in the care or safety of a child. Care proceedings are brought by local authority (LA) social care departments under section 31 of the Children Act 1989 due to concerns about significant harm, or risk of significant harm to the child attributable to care given by the parents or the child being beyond parental control. The court decides on whether or not to place the child under LA care or supervision. Of the 10,000 mothers involved with care proceedings each year in England, over 80% have their child(ren) placed in state care or in kinship care under state supervision [1, 2].

Population-based retrospective and prospective studies using linked administrative data in Wales and Canada report that young maternal age, living in the most deprived neighbourhoods, and problems such as poor mental health, substance use problems, and domestic violence are risk factors for care proceedings or children being taken into care, and often precede childbearing [3–7]. Qualitative studies in England reveal that around 2 in 5 mothers involved in recurrent care proceedings have themselves experienced child maltreatment [8]. Prospective, population-based administrative data studies in Canada and Sweeden report that mothers of children taken into care have elevated mortality compared with peers [9–11]. These findings suggest that early health interventions and intensive support for parenting might avoid court involvement entirely or mitigate the resulting health harms on parents and children.

National, whole population administrative data linking healthcare utilisation to records of family court care proceedings has not before been available for the whole of England. Pearson *et al.* (2021) linked administrative data from family court care proceedings to de-identified records from mental health services in 4 local authorities in South London, involving 3,200 mothers followed for up to 12 years [12]. Two-thirds of mothers had contact with mental health services- 80% of whom received secondary or tertiary mental health care before onset of family court care proceedings [13].

We developed an administrative database to research the upstream health determinants and outcomes of care proceedings for mothers and their children in England. We linked administrative data from Children and Family Court Advisory and Support Service (Cafcass) care proceedings and hospital records from Hospital Episode Statistics (HES) for women giving birth in an NHS hospital in England. In this study, we describe how the linked cohorts were derived and assess the match rate and indirect evidence of potential false or missed matches. We also estimate the cumulative incidence of care proceedings within 10 years of a first live birth, according to demographic subgroups, and discuss potential biases relating to linkage error. These findings are relevant to policy makers in health and family justice and will underpin applications for permissions to enable wider access to these linked data for wider research.

# Methods

## Data sources, flows and linkage

We used routinely recorded administrative data collated from NHS hospitals in England (HES) and data from public law family court care proceedings, collated by Cafcass [14, 15]. The data flows are summarised in Figure 1 and briefly described below. For more detail see Supplementary Material Sections 1–3.

### Cafcass: Public law family court care proceedings

Cafcass holds data on all public family law cases involving section 31 applications in England (Supplementary Table 1). Children in public care proceedings are automatically party to proceedings and are represented by a Cafcass guardian and their own solicitor (see definitions relevant to Cafcass in Supplementary Material Section 1). Cafcass was established on 1$^{st}$ April 2001 and de-identified data recording cases have been usable from 1$^{st}$ April 2007 [14]. Data associated with each case includes dates of hearings, number and ages of children involved in the application and age of adults considered as carers, and thereby, party to the case.

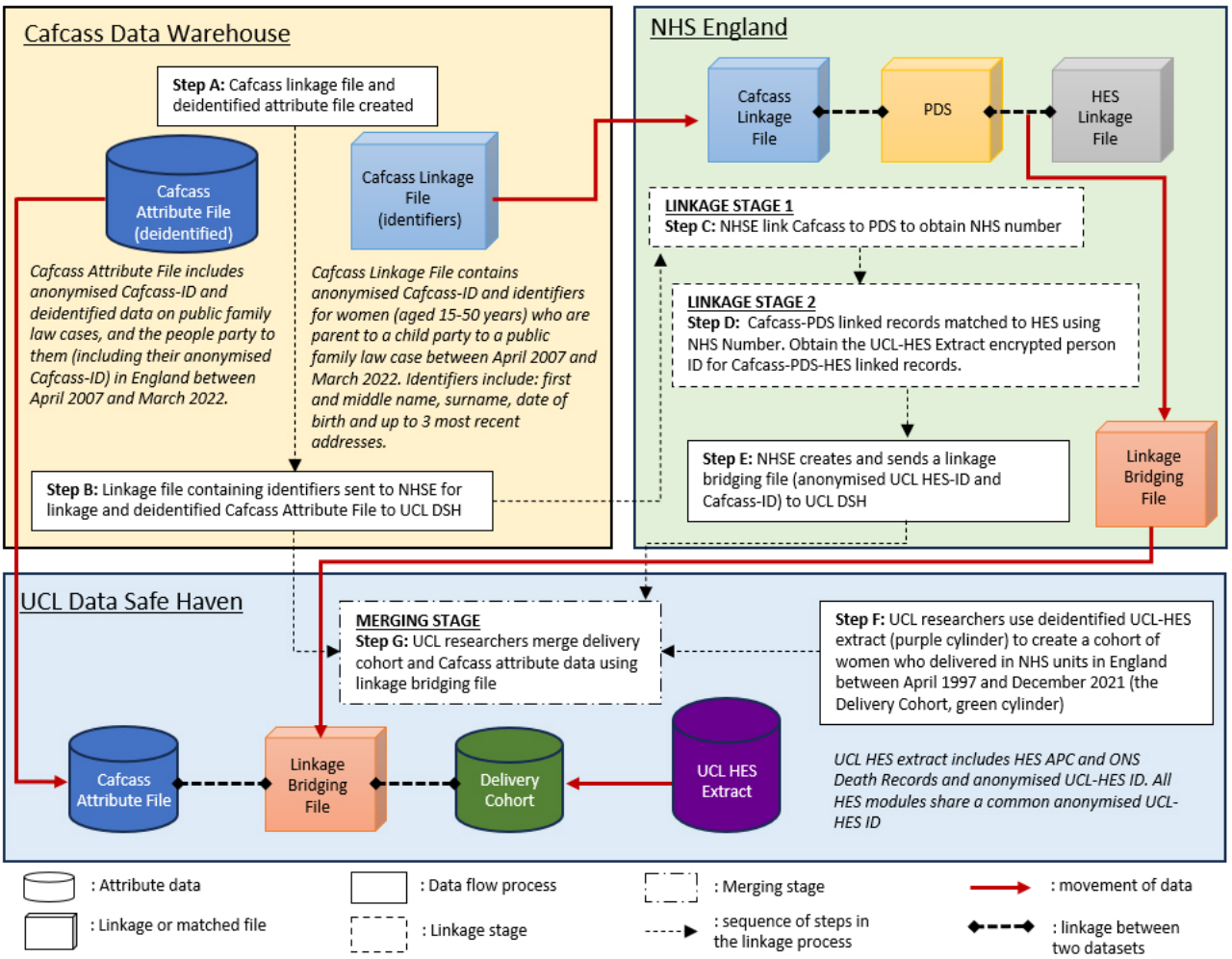
Linkage was conducted following the separation principle whereby identifiers were transferred from Cafcass to NHS England (NHSE) and only attribute data were transferred to the UCL Data Safe Haven [16]. Figure 1 Step A shows the creation of a Cafcass extract within the Cafcass secure case management system for mothers aged 15-50 at the start of their first case (1$^{st}$ April 2007 to 31$^{st}$ December 2021) and classified as a parent in section 31 proceedings. Cases were deduplicated probabilistically within Cafcass, using name, date of birth and sex before transfer to NHSE. Step B shows the transfer of the Cafcass linkage file from Cafcass to NHSE, which contained first name, middle name (where available), surname, date of birth, sex (all female) and address ($\leq$3 most recent addresses). Identifiers available in Cafcass, the personal demographic service (PDS) and HES are listed in Supplementary Material Section 2.1, Table 2. Separately, a de-identified Cafcass attribute file containing attribute data on family court care proceedings were transferred from Cafcass to the UCL Data Safe Haven. This process is known as the separation principle. A project-specific anonymised Cafcass-ID was attached to each record in the attribute and linkage files.

### NHS England (NHSE): Personal Demographic Service (PDS)

NHSE linked the Cafcass linkage file to PDS. PDS contains a full chronology of identifiers from 2004 onwards for all people who are registered with an NHS GP or referred to or attend NHS care in England [17].

During linkage stage 1 (Step C), NHSE linked the Cafcass linkage file to PDS using two automated processes: the alphanumeric trace and algorithmic trace (Supplementary Material Section 2.2) [18]. The alphanumeric trace is a deterministic link using name, date of birth, sex and address. If Cafcass records were not linked to PDS during the

Figure 1: Data flow and linkage process between administrative datasets from Cafcass, the Personal Demographic Service and Hospital Episode Statistics




Notes: Dark shading indicated de-identified and light shading identified data. Cafcass linkage data refers only to identities associated with mothers aged 15-50 years at start of a section 31 public family court case in England between 1st April 2007 and 31st December 2021. PDS = Personal Demographic Service, HES = Hospital Episode Statistics, UCL = University College London, DSH = Data Safe Haven.

alphanumeric trace, the probabilistic algorithmic trace was used. The algorithmic trace block linked the remaining records using combinations of name, date of birth, sex and postcode, and match options were scored. Each PDS candidate was scored based on the similarity of features from the query record. The score is calculated from the average of similarity scores of given and family names, date of birth (including partial date of birth), postcode (including partial and $\leq$3 historical postcodes) and sex. The highest scoring candidate PDS record was chosen as the matching record. If no candidates were found, or >1 PDS record had the highest score, then no record was returned. Manual review of links within NHSE was not possible. The NHS number was obtained from PDS for Cafcass-PDS linked records.

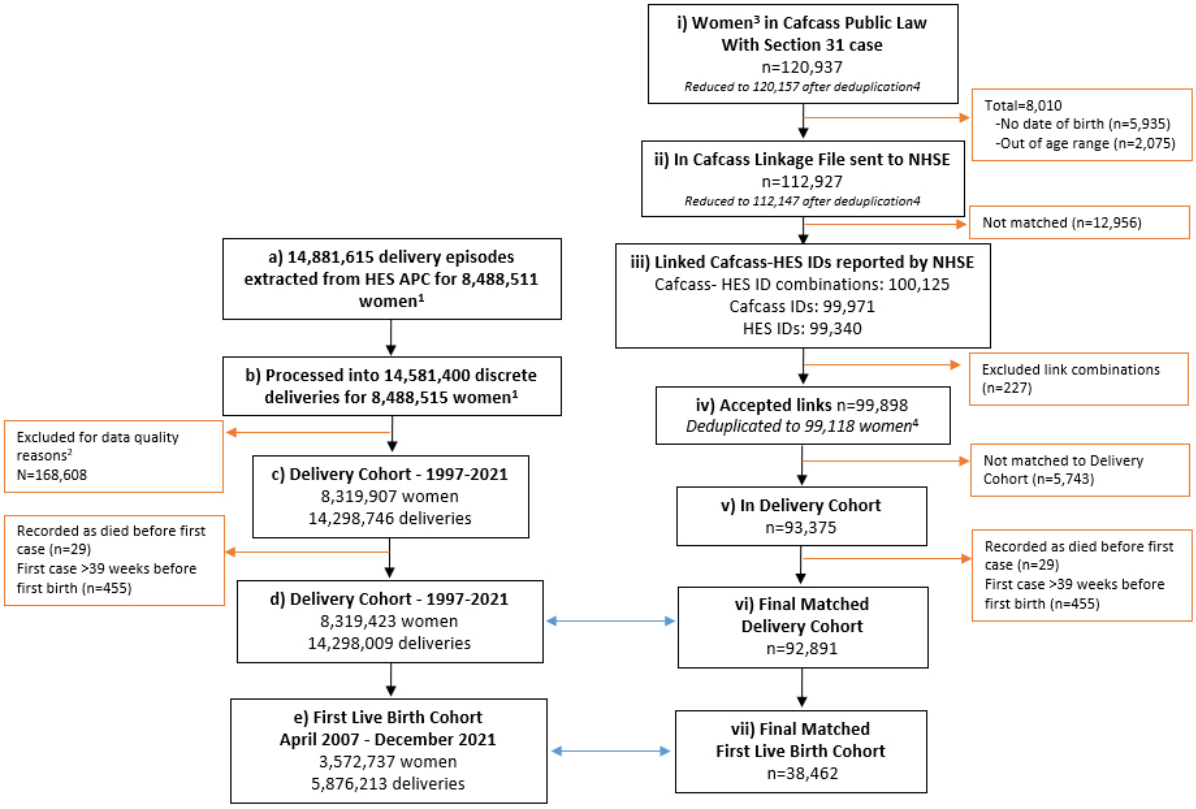
During linkage stage 2 (Step D), Cafcass-PDS linked identifiers were matched to HES identifiers using the NHS number from stage 1 and the NHS number was replaced with the project specific HES-ID for UCL. NHSE then securely transferred a linkage bridging file, containing only the Cafcass-ID and the UCL-HES-ID to the UCL Data Safe Haven (Step E). The Cafcass identifier linkage file was destroyed six months after NHSE first received it.

### UCL Hospital Episode Statistics (HES): Extract for mothers with a delivery admission

We used an extract of HES, supplied by NHSE that contained all Admitted Patient Care (APC) and any death registration records for mothers with a record of a live or still birth in a NHS unit in England between 1st April 1997 and 31st December 2021 [15, 19]. Derivation of the delivery cohort (Figure 2 boxes a to d) is described in detail in Supplementary Materials Section 3). We excluded mothers who were not recorded as resident in England and those aged less than 15 years or older than 50 years at first delivery. All hospital admissions for the same woman were linked over time using the anonymised UCL-HES-ID provided by NHSE, specifically encrypted for our data sharing agreement. Data were analysed in the UCL data safe haven.

Linkage of the deidentified Cafcass attribute file to the HES records was done in the UCL Data Safe Haven using the linkage bridging file (Step F). NHSE did not provide any information on which records linked at stage 1 but not at stage 2 (i.e., they only provided information on those who linked at both stages).

Figure 2: Results of data extraction from Cafcass and HES delivery records and linkage of Cafcass to HES, the Delivery Cohort and the First Live Birth Cohort




*Notes*

Orange box represents exclusion from Delivery or Cafcass data at each stage.

Light blue line represents number of mothers within Cafcass who linked to the Delivery and First Live Birth Cohorts.

1-Includes delivery episodes for women 12-14 years old at delivery. Women under 15 years old at first delivery were excluded following processing of delivery episodes (steps b->c).

2-see Supplementary Material Section 3.4

3-female, party to public family law case and parent of child involved.

4-Deduplicated using HES ID on accepted matches during step iv. Deduplicated when 1 HES ID matched to multiple mothers in Cafcass, following which the Cafcass records of these mothers were merged.

## Study populations

We used the HES extract to create two study populations. First, we derived a delivery cohort of all mothers residing in England who had a delivery record for a live or still born baby recorded in HES between 1st April 1997 and 31st December 2021 linked to Cafcass data (Figure 1 step F, Figure 2-a to d). The delivery cohort was used to determine the match rate and evidence of false positive matches. Second, we created a subset of mothers in the delivery cohort with a first live birth (FLB), referred to as the FLB cohort, between 1st April 2007 and 31 December 2021 (Figure 2-e). The FLB cohort included mothers involved in care proceedings and their children born from 1st April 2007 in England (Figure 2-vii). We used the FLB cohort to estimate the cumulative incidence of Cafcass proceedings after a first birth (Figure 2-e).

The process of creating a delivery cohort of women who gave birth in the NHS, including sequential delivery episodes, in England between 1st April 1997 and 31st December 2021 was conducted in four stages (Supplementary Material Section 3). In stage 1, we extracted delivery episodes from HES APC based on ICD 10 diagnosis and OPCS procedure codes and maternity tail (MT) variables [19], identified the first delivery and estimated delivery date for each woman.

In stage 2, we identified subsequent deliveries that were at least 24 weeks after the first delivery. We repeated the process to identify deliveries and estimated date of delivery until no further delivery episodes were available for each woman.

In stage 3, we assigned birth status (live or stillborn) and identified multiple births using diagnostic and maternity tail information (Supplementary Material Section 3.3). We adjusted the sequence of first, second and subsequent births by comparing the sequence of births in HES APC and the maternity tail variable indicating the number of registrable pregnancies prior to the delivery. Finally, exclusion criteria were applied to the cohort to maintain data quality, removing 217,835 mothers (Figure 2-b to c). A full list of exclusion criteria are provided in Supplementary Material Section 3.4 but include under 15 years at first delivery, resident outside of England, poor data quality from specific providers and years, delivery episodes <24 weeks apart and maternal deaths recorded prior to or between deliveries.

We identified first-time mothers to derive the FLB cohort from 1st April 2007 until 31st December 2021 (Figure 2-e; Supplementary Material Section 3.5).

To assess representativeness of the delivery cohort to all births in England, we compared annual aggregate numbers and prevalences from the Delivery Cohort with birth registrations published by the Office for National Statistics (ONS) for England for: annual number of deliveries (to provide delivery cohort coverage estimates); maternal age-group and deprivation quintile at delivery at 5 year intervals between 2000 and 2020; annual proportion of deliveries with multiple births and stillbirths; and the annual proportion of all deliveries born to primiparous women (first-time mothers (Supplementary Material Section 4) [20].

#### Evaluation of linkage quality

Linkage between de-identified Cafcass attribute file and the delivery cohort occurred at Step G (Figure 1) using the linkage bridging file. Figure 2 shows the number of unique mothers with care proceedings in Cafcass attributed records transferred to UCL DSH at Step B (Figure 2-i), the number of mothers sent to NHSE for linkage (Figure 2-ii) and the number of linked cafcass ids included in the returned linkage bridging file (Figure 2-iii).

Cafcass-HES matches in the returned bridging file were assessed. Results for mothers who linked to PDS but not to a HES record were not provided by NHSE.

We determined the overall match rate by dividing the numerator of mothers recorded in Cafcass who linked to the HES delivery cohort (1997–2021, Figure 2-vi) by the number of all mothers recorded in Cafcass and eligible for linkage (Figure 2-iii).

To evaluate potential bias resulting from missed links between Cafcass and the HES delivery cohort, we described demographic characteristics of mothers recorded in Cafcass by whether they linked to the delivery cohort or not.

Univariate and multivariable logistic regression, adjusted for ethnicity, age, number of children, year and region at first case, number of cases between 2007 and 2021 and postcode availability was used to assess predictors of linkage to a delivery record among mothers eligible to link to HES (had date of birth and between 15–50 at first case). Unlinked records could be interpreted as missed links, but could also include duplicate records of mothers in Cafcass (with one record having poorer identifiers) or mothers who linked to more than one HES ID. We could not distinguish duplicates as we did not have access to identifiable records held by Cafcass. Unlinked mothers could also have given birth outside England or prior to 1997.

To provide indirect evidence of potential false links, or erroneous recordings, we explored disagreement between characteristics that were recorded in both HES and Cafcass among mothers with a first live birth from 1st April 2007 who linked to Cafcass. Common variables in HES and Cafcass were ethnicity, region, number of children, age at first delivery (HES)/oldest child(Cafcass) and year of birth (estimated in Cafcass using year of first case and age at case start and allowing for $\pm 1$ year difference). For each variable, HES and Cafcass could either agree, disagree or the information was missing from one or both datasets. Frequency of combinations were reviewed and overall disagreement scores calculated (for each variable: 0 if agree or missing, 1 if disagree). Disagreement provided indirect evidence of a potential false match (although we could not distinguish between errors in data recording and linkage errors).

### Estimation of the cumulative incidence of care proceedings at 10 years

We use Kaplan-Meier analyses to estimate the cumulative incidence, and 95% confidence interval (CI), of a first care proceedings at 5 and 10 years after first live birth (using Stata/MP Version 18). Follow-up started at first live birth and ended at whichever came first; start of first care proceeding, the date 10 years after first live birth, 31st December 2021, or death (using linked death registration or death recorded as discharge method in HES). Cumulative incidence at 10 years was estimated for the following population subgroups: maternal age-group, ethnic group, quintile of area-based index of multiple deprivation (IMD) a measure of the concentration of deprived households within a small area (lower super output area, which contains an average of 1,500 individuals), year of first live birth and health region of first delivery [21].

### Sensitivity analysis to addressing underestimation of cumulative incidence estimates

Match rates demonstrate that underestimation of the cumulative incidence of care proceedings is likely. The extent to which incidence of care proceedings within 10 years of a FLB was underestimated due to missed matches was also assessed quantitatively using the records of mothers who did not link to HES. Among unlinked mothers recorded only in Cafcass, year of FLB was estimated using maternal age at oldest child and first case start (where available). We excluded mothers with an estimated FLB before 2007. Under scenario 1, the least conservative estimate, all remaining mothers were added to the crude numerator of mothers who experienced care proceedings within 10 years of a FLB. Under scenario 2, only mothers with a child less than 2 years at first case were added to the numerator, as we assumed that births close in time to care proceedings were more likely to have been in England. The percent increase in crude incidence from scenario 1 and 2 was applied to Kaplan-Meier estimates of cumulative incidence. These scenarios assume there are no duplicate mothers in the unlinked pool, that all children associated with the case were delivered within the NHS in England, all mothers without a date of birth were 15-50 years old at case start and that unlinked mothers would contribute similar follow-up time to linked mothers.

## Results

### Data flows and linkage

Between 1st April 2007 and 31st December 2021, 120,937 mothers were party to care proceedings involving one or more children (Figure 2-i). The number of unique mothers eligible for linkage was reduced to 112,927 (Figure 2-ii) because mother's date of birth was not recorded in 4.9% (n = 5,935) and maternal age was outside age 15-50 years

(1.7%; n = 2,075). Recording of date of birth varied by ethnic background, number of care proceedings and region- with London (9.2%) and the East of England (6.8%) having the highest proportions of mothers missing date of birth (Supplementary Table 1). Completeness of date of birth improved over time, from 90.2% in 2007-10 to 97.6% in 2019–21. An additional 1.7% (n = 2,075) were excluded as they were not aged between 15 and 50 years at start of first case. These 112,927 mothers (93.4%) were included in the Cafcass Linkage File and were sent to NHSE.

### Study populations

Between 1st April 1997 and 31st December 2021, 14,881,615 delivery episodes in 8,488,511 women aged 12–50 years old were extracted from HES (Figure 2-a, see Supplementary Material Section 3). Processing steps identified 14,581,400 separate deliveries in these women. Overall, 168,608 women (equivalent to 282,654 deliveries) were excluded from the cohort (see Supplementary Material Section 3.4), leaving a total of 8,319,907 women aged 15 to 50 years with 14,298,746 deliveries.

For the 1998-2021 period, the Delivery Cohort was estimated to contain 92.9% of all births registered in England by the Office for National Statistics, increasing from 87.0% in 1998 to 95.2% in 2013, then falling to 91.7% in 2021. Further comparison of the delivery cohort to ONS data on birth registrations for England is reported in Supplementary Material Section 4.

The FLB cohort comprised a subset of 3,572,737 mothers in the delivery cohort with a first recorded live birth from 1st April 2007 (Figure 2-e). There were a further 2,303,476 deliveries (total 5,876,213) for these mothers up to end-2021. In the FLB cohort, 1.4% of first time and 1.3% of subsequent deliveries included more than 1 infant and 0.3% of subsequent deliveries included at least 1 still born infant.

### Evaluation of linkage quality

The Linkage Bridging File transferred from NHSE to UCL for analysis (Figure 1 Step E, Figure 2-iii) contained 100,125 unique Cafcass-HES ID combinations, with 88.5% of Cafcass records (99,971/112,927) matching to at least one HES record (99,340 HESIDs). Where mothers linked to 2 or more HES IDs, the most common link was accepted ($\leq$3 possible), otherwise the matches were discarded (77 combinations accepted, 227 rejected).

Amongst the 99,898 accepted matches (Figure 2-iv), where 2 or more mothers linked to the same HES ID (n = 1,708) the care proceedings records were merged into a single ID (n = 929), thus removing duplicates missed prior to linkage and resulting in 99,118 mothers who matched to HES, and a final HES match rate of 88.4% (99,118/112,147).

Figure 2 shows that 5,743 mothers recorded in Cafcass were linked to a HES record but not to a delivery record and were excluded. A further 455 mothers were removed from the delivery cohort as their first care proceeding started more than 39 weeks before their first liveborn delivery recorded in HES, and 29 due to date of death recorded as before to case start. This left 92,891 (82.8% of 112,147) mothers with care proceedings linked to a delivery record between 1997 and 2021 (Figure 2-vi). Of these, 38,462 mothers had a first live birth between 1st April 2007 and 31st December 2021 (34.3% of 112,147) and were followed up in the FLB cohort of 3,572,737 mothers (Figure 2-vii).

### Characteristics of mothers in Cafcass according to linkage status

Of the 112,147 mothers recorded in Cafcass with identifiers for linkage, 13,029 (11.6%) did not link to any HES record, 6,227 (5.6%) linked to a HES record but the final Delivery Cohort, and 92,891 (82.8%) linked to the Delivery Cohort (Table 1 and Figure 2). Mothers from an ethnic background classified as black or other, aged over 40 years at first case, with a first case between 2007–2010 and whose first case was in London or the East of England, were least likely to link to the Delivery Cohort.

Additionally, an adjusted logistic regression analysis to summarise associations between characteristics of mothers recorded in Cafcass and linkage to the Delivery Cohort showed mothers with a postcode recorded (92.0%; 8,939/112,147) were 1.77 times (95% Confidence Interval [CI]: 1.69-1.87) more likely than mothers without a postcode in Cafcass to match to a HES delivery record (Supplementary Material Section 2.2).

### Disagreement between common variables in HES and Cafcass

Missingness was low ($\leq$0.5%) for all variables other than ethnicity (where 22.6% of match combinations had ethnicity missing from either data source). Of the 38,462 mothers recorded in Cafcass linked to the FLB Cohort, 74.6% disagreed on no variables (Supplementary Table 8a). Disagreement was highest for the number of children at case start (17.2%, n = 6,608), followed by age at first child (10.7%, n = 4,126), ethnicity (4.2%, n = 1,634), region (2.7%, n = 1,051) and maternal year of birth (1.0%, n = 383) and Supplementary Table 8b shows the 5 most common patterns of disagreement.

### Cumulative Incidence of care proceedings

The FLB cohort contained 3,572,737 mothers, with 5,876,213 deliveries, and 1.1% of mothers (38,462/3,572,737) had a Cafcass care proceeding between 1st April 2007 and 31st December 2021 (Table 2). Overall, the majority of deliveries were singletons (98.6%; n = 5,794,598), with 1.4% (n = 81,615) recorded as a multiple, and 0.15% (n = 8,701) were recorded as having at least 1 still born infant and in <0.03% (n = 1,372) the outcome was uncertain. 6,655 (0.2%) of mothers died within 10 years of their first live birth.

Compared to mothers not in care proceedings, mothers subject to care proceedings were younger at first delivery (78.4% vs 28.8% under 25 years) and more deprived (72.0% vs 47.0% residing in the two most deprived quintiles).

The cumulative incidence of a care proceeding by 10 years after a first live birth was 1.31% (95% CI: 1.29–1.32%) (Table 2; Figure 3). Cumulative incidence was highest in mothers who were aged 15–19 years (6.79%, 95% CI: 6.69–6.89%) or 20–24 years (1.89%, 95% CI: 1.85–1.92%) at first delivery, most deprived at first delivery (2.47%, 95% CI: 2.49–2.51%) and of mixed ethnicity (1.93%, 95% CI: 1.80–2.07%). Sensitivity analysis (Supplementary Material Section 6), using

Table 1: Characteristics of mothers recorded in Cafcass care proceedings between 1st April 2007 and 31st December 2021 in England according to linkage status: linked to a UCL-HES extract record and linked to a delivery cohort record

| Characteristics recorded in Cafcass | Mothers | %[1] | Linked to UCL HES | | | Linked to Delivery Cohort | | |
|---|---|---|---|---|---|---|---|---|
| | | | n | % | MR[1,2] | n | % | MR[1,2] |
| *Total* | 120,157 | | 9,118 | | | 92,891 | | |
| Not available for linkage[3] | 8,010 | | | | | | | |
| Available for linkage | 112,147 | 100 | 99,118 | 100 | 88.4 | 92,891 | 100 | 82.8 |
| Ethnicity[4] | | | | | | | | |
| White | 70,470 | 62.8 | 63,385 | 63.9 | 89.9 | 60,251 | 64.9 | 85.5 |
| Black | 4,487 | 4.0 | 3,573 | 3.6 | 79.6 | 3,127 | 3.4 | 69.7 |
| Asian | 3,116 | 2.8 | 2,603 | 2.6 | 83.5 | 2,340 | 2.5 | 75.1 |
| Mixed | 3,067 | 2.7 | 2,642 | 2.7 | 86.1 | 2,470 | 2.7 | 80.5 |
| Other | 1,000 | 0.9 | 797 | 0.8 | 79.7 | 647 | 0.7 | 64.7 |
| Missing | 30,007 | 25.0 | 26,118 | 26.4 | 87.0 | 24,056 | 25.9 | 80.2 |
| Number of children[4] | | | | | | | | |
| 1 | 48,091 | 42.9 | 41,863 | 42.2 | 87.0 | 38,242 | 41.2 | 79.5 |
| 2 | 29,446 | 26.3 | 26,076 | 26.3 | 88.6 | 24,706 | 26.6 | 83.9 |
| 3 | 17,769 | 15.8 | 15,876 | 16.0 | 89.3 | 15,264 | 16.4 | 85.9 |
| 4+ | 16,841 | 15.0 | 15,303 | 15.4 | 90.9 | 14,679 | 15.8 | 87.2 |
| Age at first s31 case | | | | | | | | |
| 15–19 | 11,902 | 10.6 | 10,539 | 10.6 | 88.5 | 9,978 | 10.7 | 83.8 |
| 20–24 | 22,858 | 20.4 | 20,284 | 20.5 | 88.7 | 19,551 | 21.0 | 85.5 |
| 25–29 | 22,659 | 20.2 | 20,112 | 20.3 | 88.8 | 19,341 | 20.8 | 85.4 |
| 30–34 | 22,169 | 19.8 | 19,680 | 19.9 | 88.8 | 18,635 | 20.1 | 84.1 |
| 35–39 | 17,149 | 15.3 | 15,145 | 15.3 | 88.3 | 13,931 | 15.0 | 81.2 |
| 40–44 | 10,093 | 9.0 | 8,822 | 8.9 | 87.4 | 7,703 | 8.3 | 76.3 |
| 45–50 | 5,317 | 4.7 | 4,536 | 4.6 | 85.3 | 3,752 | 4.0 | 70.6 |
| Year of first s31 case | | | | | | | | |
| 2007–2010 | 20,968 | 18.7 | 18,574 | 18.7 | 88.6 | 16,958 | 18.3 | 80.9 |
| 2011–2014 | 29,837 | 26.6 | 26,293 | 26.5 | 88.1 | 24,638 | 26.5 | 82.6 |
| 2015–2018 | 36,976 | 33.0 | 32,526 | 32.8 | 88.0 | 30,710 | 33.1 | 83.1 |
| 2019–2022 | 24,366 | 21.7 | 21,725 | 21.9 | 89.2 | 20,585 | 22.2 | 84.5 |
| Region of first case[5] | | | | | | | | |
| North East | 9,225 | 8.2 | 8,385 | 8.5 | 90.9 | 8,015 | 8.6 | 86.9 |
| North West | 18,865 | 16.8 | 16,675 | 16.8 | 88.4 | 15,755 | 17.0 | 83.5 |
| York. & Humber[6] | 13,085 | 11.7 | 11,915 | 12.0 | 91.1 | 11,250 | 12.1 | 86.0 |
| East Midlands | 8,875 | 7.9 | 7,985 | 8.1 | 90.0 | 7,535 | 8.1 | 84.9 |
| West Midlands | 12,050 | 10.7 | 10,730 | 10.8 | 89.0 | 10,110 | 10.9 | 83.9 |
| East of England | 9,635 | 8.6 | 8,545 | 8.6 | 88.7 | 7,970 | 8.6 | 82.7 |
| London | 15,435 | 13.8 | 12,785 | 12.9 | 82.8 | 11,455 | 12.3 | 74.2 |
| South East | 14,700 | 13.1 | 13,055 | 13.2 | 88.8 | 12,240 | 13.2 | 83.3 |
| South West | 9,775 | 8.7 | 8,610 | 8.7 | 88.1 | 8,155 | 8.8 | 83.4 |
| Not available | 505 | 0.5 | 430 | 0.4 | 85.1 | 405 | 0.4 | 80.2 |
| Number of s31 cases | | | | | | | | |
| 1 | 85,606 | 76.3 | 75,119 | 75.8 | 87.7 | 69,885 | 75.2 | 81.6 |
| 2 | 19,328 | 17.2 | 17,459 | 17.6 | 90.3 | 16,714 | 18.0 | 86.5 |
| 3+ | 7,213 | 6.4 | 6,540 | 6.6 | 90.7 | 6,292 | 6.8 | 87.2 |

N.B. Characteristics are based on information as recorded within Cafcass, and thus may differ to characteristics recorded in HES and presented in Table 2.
1-Of those available for linkage (max 112,147), calculated using row total as a denominator.
2-MR = Match rate.
3-due to no information on date of birth or if outside age range (15-50 years) included for linkage.
4-As recorded in Cafcass.
5-Rounded to nearest 5 in accordance with NHS England disclosure control methodology for HES.
6-Yorkshire and the Humber.

Table 2: Characteristics of mothers, aged 15-50 years, who had a live birth in an NHS service, and were resident in England between 1st April 2007 and 31st December 2021 (the First Live Birth Cohort), number and proportion with care proceedings and the cumulative incidence, and 95% Confidence Interval (CI), of care proceedings within 10 years of a first live birth

| | Mothers | | Not experienced care proceedings | | Experienced care proceedings | | | |
|---|---|---|---|---|---|---|---|---|
| | n | % | n | % of total | n | % of total | 10-year cumulative incidence | 95% Confidence interval |
| Total | 3,572,737 | 100 | 3,534,275 | 100.0 | 38,462 | 100.0 | 1.31 | 1.29–1.32 |
| Age at first delivery | | | | | | | | |
| <20 | 292,316 | 8.2 | 274,201 | 7.8 | 18,115 | 47.1 | 6.79 | 6.69–6.89 |
| 20–24 | 754,819 | 21.1 | 742,766 | 21.0 | 12,053 | 31.3 | 1.89 | 1.85–1.92 |
| 25–29 | 1,044,397 | 29.2 | 1,039,939 | 29.4 | 4,458 | 11.6 | 0.53 | 0.51–0.54 |
| 30–34 | 978,389 | 27.4 | 976,077 | 27.6 | 2,312 | 6.0 | 0.30 | 0.28–0.31 |
| 35–39 | 412,709 | 11.6 | 411,505 | 11.6 | 1,204 | 3.1 | 0.36 | 0.34–0.38 |
| 40–44 | 84,146 | 2.4 | 83,836 | 2.4 | 310 | 0.8 | | |
| 45–50 | 5,961 | 0.2 | 5,951 | 0.2 | 10 | 0.0 | | |
| Ethnicity | | | | | | | | |
| White | 2,621,788 | 73.4 | 2,588,348 | 73.2 | 33,440 | 86.9 | 1.51 | 1.49–1.53 |
| Black | 137,028 | 3.8 | 135,554 | 3.8 | 1,474 | 3.8 | 1.25 | 1.18–1.32 |
| Asian | 349,866 | 9.8 | 348,688 | 9.9 | 1,178 | 3.1 | 0.40 | 0.38–0.43 |
| Mixed | 57,181 | 1.6 | 56,275 | 1.6 | 906 | 2.4 | 1.93 | 1.80–2.07 |
| Other | 121,865 | 3.4 | 121,199 | 3.4 | 666 | 1.7 | 0.67 | 0.61–0.72 |
| Not reported | 285,009 | 8.0 | 284,211 | 8.0 | 798 | 2.1 | 0.43 | 0.40–0.46 |
| IMD Quintile[1] | | | | | | | | |
| 1-*most deprived* | 872,836 | 24.4 | 854,949 | 24.2 | 17,887 | 46.5 | 2.47 | 2.43–2.51 |
| 2 | 812,742 | 22.7 | 803,046 | 22.7 | 9,696 | 25.2 | 1.43 | 1.40–1.46 |
| 3 | 701,728 | 19.6 | 696,311 | 19.7 | 5,417 | 14.1 | 0.95 | 0.92–0.98 |
| 4 | 611,178 | 17.1 | 607,919 | 17.2 | 3,259 | 8.5 | 0.65 | 0.63–0.67 |
| 5-*least deprived* | 566,446 | 15.9 | 564,380 | 16.0 | 2,066 | 5.4 | 0.45 | 0.42–0.47 |
| Not reported | 7,807 | 0.2 | 7,670 | 0.2 | 137 | 0.4 | 1.57 | 1.32–1.88 |
| Region[1,2] | | | | | | | | |
| North East | 154,310 | 4.3 | 150,936 | 4.3 | 3,374 | 8.8 | 2.62 | 2.52–2.71 |
| North West | 439,805 | 12.3 | 433,483 | 12.3 | 6,324 | 16.4 | 1.71 | 1.67–1.76 |
| York. & Humber[3] | 324,220 | 9.1 | 319,604 | 9.0 | 4,614 | 12.0 | 1.63 | 1.58–1.68 |
| East Midlands | 250,685 | 7.0 | 247,693 | 7.0 | 2,994 | 7.8 | 1.38 | 1.33–1.44 |
| West Midlands | 342,505 | 9.6 | 338,421 | 9.6 | 4,085 | 10.6 | 1.41 | 1.37–1.46 |
| East of England | 363,315 | 10.2 | 359,996 | 10.2 | 3,318 | 8.6 | 1.09 | 1.05–1.13 |
| London | 669,390 | 18.7 | 665,132 | 18.8 | 4,258 | 11.1 | 0.75 | 0.72–0.77 |
| South East | 530,995 | 14.9 | 525,953 | 14.9 | 5,041 | 13.1 | 1.11 | 1.08–1.15 |
| South West | 318,155 | 8.9 | 314,467 | 8.9 | 3,689 | 9.6 | 1.35 | 1.30–1.40 |
| No Fixed Abode | 565 | 0.0 | 536 | 0.0 | 30 | 0.1 | 5.60 | 3.94–7.95 |
| Not reported | 178,790 | 5.0 | 178,054 | 5.0 | 735 | 1.9 | 1.36 | 1.16–1.60 |
| Year of first delivery | | | | | | | | |
| 2007[4]–2010 | 1,010,671 | 28.3 | 996,179 | 28.2 | 14,492 | 37.7 | 1.20 | 1.18–1.22 |
| 2011–2014 | 993,652 | 27.8 | 981,664 | 27.8 | 11,988 | 31.2 | 1.28 | 1.26–1.31 |
| 2015–2018 | 928,000 | 26.0 | 919,436 | 26.0 | 8,564 | 22.3 | | |
| 2019–2022 | 640,414 | 17.9 | 636,996 | 18.0 | 3,418 | 8.9 | | |
| Multiples[1] | | | | | | | | |
| Singleton | 3,521,210 | 98.6 | 3,483,141 | 98.6 | 38,069 | 99.0 | 1.32 | 1.30–1.33 |
| Multiples | 51,527 | 1.4 | 51,134 | 1.4 | 393 | 1.0 | 0.88 | 0.79–0.97 |
| Birth Status[1] | | | | | | | | |
| All live born | 3,571,397 | 100.0 | 3,532,946 | 100.0 | 38,451 | 100.0 | | |
| Some live born, some still born | 1,146 | 0.0 | 1,137 | 0.0 | c. | 0.0 | | |
| Uncertain | 194 | 0.0 | 192 | 0.0 | c. | 0.0 | | |

N.B. Characteristics are based on information as recorded within HES and the delivery cohort, and thus may differ to characteristics recorded in Cafcass and presented in Table 1.

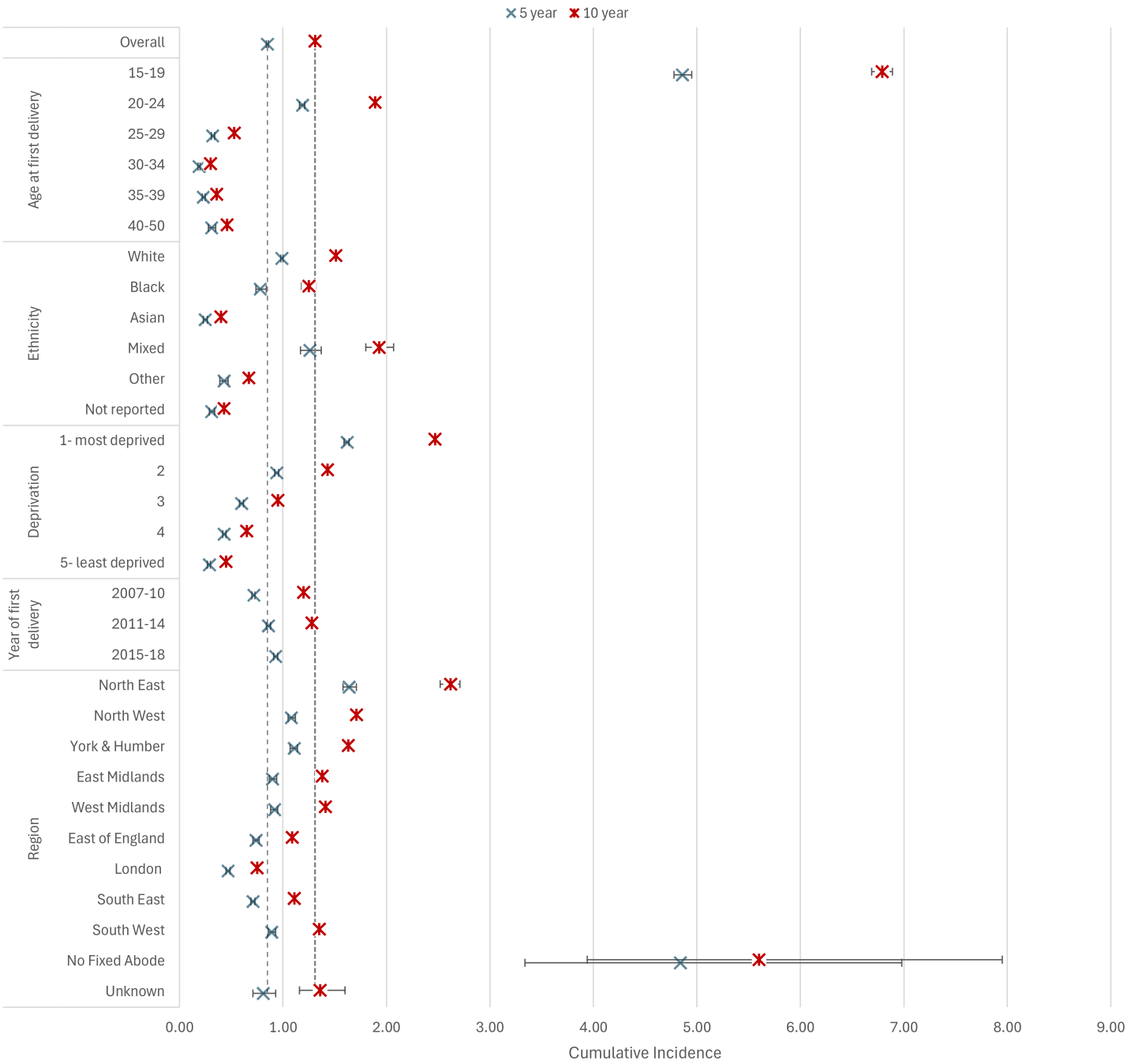
1-based on first delivery between 1st April 2007 and 31stDecember 2021.

2-Rounded to nearest 5 in accordance with NHS England disclosure control methodology for HES.

3-Yorkshire and the Humber.

4-from 1stApril 2007.

Figure 3: Kaplan-Meier estimates of cumulative incidence of care proceedings at 5 and 10 years after a first live birth, with 95% confidence intervals, between 1st April 2007 and 31st December 2021




data from unlinked mothers, increased estimated cumulative incidence up to 1.7%.

## Discussion

We linked 88.4% of mothers recorded in Cafcass care proceedings to a HES admission, and 82.4% to a HES delivery record. Mothers residing in London, East of England, older age and reported as black or other ethnic background were least likely to link to a HES delivery record. Amongst mothers with care proceedings who linked to the first live birth cohort, 74% had Cafcass and HES records agree on five compared variables, but disagreement on number or age of children was present for 26%. The cumulative incidence of Cafcass care proceedings within 10 years of a first live birth was 1.3% for all mothers, and higher for mothers aged less than 20 years at their first live birth (6.9%), most deprived (2.50%) and of mixed ethnicity (1.96%).

### Results in context

For the first time, national court and hospital records for England were linked and our match rate is comparable

to that reported for Welsh data [22]. Johnson *et al.* linked Welsh family court and electronic healthcare records, including general practice records, via the Secure Anonymised Information Linkage Databank [SAIL] for all people associated with cases, and achieved a match rate of 74% [22]. In contrast, in Sweden and Canada it is possible to link across multiple databases deterministically using unique national identification numbers [9, 23]. We found match rates vary by region, maternal age and ethnicity, but similar figures are not available in Welsh-linked data. These differences will impact cumulative incidence estimates. A key driver of differences in match rates is the variability in availability of date of birth and postcode, both of which were lowest in London and the East of England. This may be because London is more ethnically diverse than other regions, a higher proportion of mothers in London may have delivered outside of England, thus not contained in the delivery cohort, or the population of London is more mobile, meaning postcode data in Cafcass may not be up to date. The higher match rate to HES via PDS of 88% highlights the potential for public law family court care proceedings data to be linked to a range of NHSE owned healthcare datasets. Of particular interest are the mental health, maternity and emergency care data services. As data linkage of Cafcass to administrative datasets continue, the data quality, particularly of linkage identifiers remains important and may require improvements.

We found deprived, younger mothers were more likely to experience care proceedings within the first 10 years of parenthood, corresponding with previous, descriptive analyses from England and Wales care proceedings data [24]. To date, English estimates have used unlinked data or Children Looked After returns. Using these data, it has been estimated that between 1.2% to 1.8% of children enter out-of-home care by 10 years, increasing to 3.3% by 18 years old [25–27]. Our estimate of 1.3% of mothers experiencing care proceedings within 10 years of a FLB, and an upper estimate of 1.7% under sensitivity scenario 1, aligns with these previous estimates. Whilst strategies to reduce teenage pregnancy rates have resulted in declining rates, teenage mothers remain at greatest risk of care proceedings [28]. Young mothers require early, sustained, co-ordinated, multiagency support to improve the outcomes of both mother and child(ren) [29]. Following assessment, support may include help to continue with education, return to work, or to develop adequate networks and avoid social isolation, as well as addressing health needs [30, 31]. However, investment in early intervention services, such as Sure Start children's centres, has fallen by 46% since 2010–11 [32]. Our results show the utility of data linkage of health and court data, and in the future, exploration of up-stream health profiles will provide further insights into predictors of care proceedings and direction of preventive measures.

## Strengths and limitations

Strengths include use of whole of England data, with up to 24 years of follow up. This allowed us to derive a FLB cohort so that all mothers can be followed from the same point in their reproductive life course. A limitation is that we could not exclude left censoring, by which we mean births that occurred outside the NHS in England, prior to 1$^{st}$ April 1997, or were not linked to the woman's records from 2007 onwards. As longer look-back periods become available and the data quality of the maternity tail improves, the impact of this censoring will be reduced. Knight *et al.* (2013), referenced unpublished data stating that 90% of mothers have their second infant within 7 years of the first, which is similar to the 86.5-90.2% we estimate based on early deliveries in our cohort [33]. This means that from 2005, less than 10% of mothers will be wrongly recorded as primiparous from cohort processing alone, a figure further reduced as 66.1% of mothers records had data on the number of previous registrable pregnancies. In 2021, 29% of live births were to non-UK born women, some of whom will be highly mobile [34]. The Delivery Cohort is also influenced by right censoring, particulary if women leave England.

The drivers of differences in match rates are discussed earlier, but the proportion of mothers who should have matched to the FLB cohort is difficult to ascertain through Cafcass data alone. We estimate, through sensitivity scenario 1, cumulative incidence could be as high as 1.7%. However, we have assumed all children associated with these cases were born in NHS maternity units in England and are not aware of data to verify this. Additionally, the records of mothers without date of birth are more likely to be unmatched, duplicate records.

Due to the data linkage process of NHE, we were unable to review match quality with identifiable data or know whether unlinked mothers did not match to any PDS records or matched to multiple PDS records with the same match quality, and so not provided in the linkage bridging file. Where mothers did match and FLB cohort and Cafcass data was compared, <0.4% of records disagreed on 4 or more variables. The most common disagreement was for age at first child and number of children, which may occur if care proceedings do not include all children and/or some children do not live in the family home (ie live with another family member).

## Conclusion

We have demonstrated successful linkage of court, health and delivery records in England, although improved identifier quality would reduce linkage biases for vulnerable groups and between regions. Using the resulting data, we estimate one in 15 mothers younger than 20 years at their first birth were involved in care proceedings within 10 years. In the future, this data will provide opportunities to better understand the health needs of mothers prior to, and following, care proceedings. The deployment of appropriate healthcare services could reduce the need for children to be removed from the family home and improve the health and wellbeing of mothers before and following child removal.

# Acknowledgments

We are grateful to the Children and Family Court Advisory and Support Service (Cafcass), for providing extracts of their case management data to establish this linkage, and the patients, their families, and NHS staff for their ongoing contribution to research. This work uses data provided by patients and collected by the National Health Service as part

of their care and support. Permission to use de-identified data from Hospital Episode Statistics was granted by NHS Digital (DARS-NIC-196263).

This project was funded by the Nuffield Foundation (grant number: JUS/FR-000020122). RG, LW and MJ were (in part) supported by the National Institute for Health and Care Research (NIHR) Children and Families Policy Research Unit (PR-PRU-1217-21301); RG by a NIHR Senior Investigator award and by Health Data Research UK (HDRUK2023.0029); GI, MJ, QF, KH, LW and RG by the NIHR GOSH Biomedical Research Centre, and KH, RG and QF by ADR UK (Administrative Data Research UK), an Economic and Social Research Council (part of UK Research and Innovation) programme (ES/V000977/1, ES/X000427/1 and ES/X003663/1). CG is supported by the Economic and Social Research Council (ESRC) UBEL Doctoral Training Programme (ES/P000592/1).

The views expressed are those of the author(s) and not necessarily those of the Nuffield Foundation, NIHR, the Department of Health and Social Care, ADR UK or ESRC.

## Data availability

Hospital Episode Statistics data and was provided within the terms of a data-sharing agreement (DARS-NIC-196263-J9Q7Z-v1.4) to the researchers by NHS England. The data do not belong to the authors and may not be shared by the authors, except in aggregate form for publication. The data is provided by patients and collected by the NHS as part of their care and support. Data can be obtained by submitting a data request through the NHS England Data Access Request Service and Cafcass.

## Statement on conflicts of interest

Nothing to declare

## Ethics statement

Ethical approval was obtained from the UCL Research Ethics Committee (reference: 19/LO/0103), and the Cafcass Research Governance Committee to work with the Cafcass data. The Health Research Authority Confidentiality Advisory Group granted exemption from s251 to enable linkage between identifiers supplied by Cafcass to NHS identifiers by NHS England. Data sharing agreements are in place to use de-identified attribute data from Cafcass and NHS England (DARS-NIC-196263-J9Q7Z-v1.4) in the UCL Data Safe Haven (privacy notice).

## Funding

This work was supported by the Nuffield Foundation [grant number JUS/FR-000020122]. The views expressed are those of the author(s) and not necessarily those of Cafcass or the Nuffield Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## References

1. Broadhurst K, Alrouh B, Yeend E, Harwin J, Shaw M, Pilling M, *et al*. Connecting Events in Time to Identify a Hidden Population: Birth Mothers and Their Children in Recurrent Care Proceedings in England. Brit J Soc Work. 2015;45(8):2241–60. https://doi.org/10.1093/bjsw/bcv130

2. Cafcass. Cafcass Official Demand Statistics. 2023. https://www.cafcass.gov.uk/about-us/our-data.

3. Wall-Wieler E, Roos LL, Brownell M, Nickel NC, Chateau D. Predictors of having a first child taken into care at birth: A population-based retrospective cohort study. Child Abuse Neglect. 2018;76:1–9. https://doi.org/10.1016/j.chiabu.2017.09.033

4. Griffiths LJ, Johnson RD, Broadhurst K, Bedston S, Cusworth L, Alrouh B, *et al*. Maternal health, pregnancy and birth outcomes for women involved in care proceedings in Wales: a linked data study. Bmc Pregnancy Childb. 2020;20(1). https://doi.org/10.1186/s12884-020-03370-4

5. Johnson RD, North L, Alrouh B, John A, Jones K, Akbari A, *et al*. A population level study into health vulnerabilities of mothers and fathers involved in public law care proceedings in Wales, UK between 2011 and 2019. International Journal of Population Data Science (Ijpds). 2022;7(1). https://doi.org/10.23889/ijpds.v7i1.1723

6. Johnson R, Alrouh B, Broadhurst K, Ford D, John A, Jones K, *et al*. Health vulnerabilities of parents in care proceedings in Wales. London: Nuffield Family Justice Observatory; 2021. https://www.nuffieldfjo.org.uk/resource/health-vulnerabilities-parents-care-proceedings.

7. Johnson R, Griffiths LJ, Cowley LE, Broadhurst K, Bailey R. Risk Factors Associated With Primary Care-Reported Domestic Violence for Women Involved in Family Law Care Proceedings: Data Linkage Observational Study. J Med Internet Res. 2023;25. https://doi.org/10.2196/42375

8. Broadhurst K, Mason C, Bedston S, Alrouh B, Morriss L, McQuarrie T, et al. Vulnerable Birth Mothers and Recurrent Care Proceeding. Lancaster University; 2017. https://www.nuffieldfoundation.org/sites/default/files/files/rc-final-summary-report-v1_a6.pdf.

9. Wall-Wieler E, Vinnerljung B, Liu C, Roos LL, Hjern A. Avoidable mortality among parents whose children were placed in care in Sweden: a population-based study. J Epidemiol Commun H. 2018;72(12):1091–8. https://doi.org/10.1136/jech-2018-210672

10. Wall-Wieler E, Roos LL, Nickel NC, Chateau D, Brownell M. Mortality Among Mothers Whose Children Were Taken Into Care by Child Protection Services: A Discordant Sibling Analysis. Am J Epidemiol. 2018;187(6):1182-8.https://doi.org/10.1093/aje/kwy062

11. Wall-Wieler E, Roos LL, Brownell M, Nickel N, Chateau D, Singal D. Suicide Attempts and Completions among Mothers Whose Children Were Taken into Care by Child Protection Services: A Cohort Study Using Linkable Administrative Data. Can J Psychiat. 2018;63(3):170–7. https://doi.org/10.1177/0706743717741058

12. Pearson RJ, Grant C, Wijlaars L, Finch E, Bedston S, Broadhurst K, Gilbert R. Mental health service use among mothers involved in public family law proceedings: linked data cohort study in South London 2007–2019. Social Psychiatry and Psychiatric Epidemiology. 2022;57(10):2097–108. https://doi.org/10.1007/s00127-022-02221-1

13. Pearson RJ. Using longitudinal administrative data to characterise mental health problems and substance misuse among women whose children enter care in England: University College London; 2021. https://discovery.ucl.ac.uk/id/eprint/10139871/.

14. Bedston SJ, Pearson RJ, Jay MA, Broadhurst K, Gilbert R, Wijlaars L. Data Resource: Children and Family Court Advisory and Support Service (Cafcass) public family law administrative records in England. International Journal of Population Data Science. 2020;5(1). https://doi.org/10.23889/ijpds.v5i1.1159

15. 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;46(4):1093-i. https://doi.org/10.1093/ije/dyx015

16. Gilbert R, Lafferty R, Hagger-Johnson G, Harron K, Zhang LC, Smith P, *et al*. GUILD: GUidance for Information about Linking Data sets. J Public Health-Uk. 2018;40(1):191–8. https://doi.org/10.1093/pubmed/fdx037

17. NHS England. Personal Demographics Service 2023 [Available from: https://digital.nhs.uk/services/personal-demographics-service.

18. NHS Digital. The Person_ID handbook for HES users. 2021. https://digital.nhs.uk/services/personal-demographics-service/master-person-service#user-guidance.

19. 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;11(10):e0164667. https://doi.org/10.1371/journal.pone.0164667

20. Office for National Statistics. Number of maternities by Index of Multiple Deprivation (IMD), age of mother, type of outcome and previous live born children 2023 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/maternities/adhocs/1637numberofmaternitiesbyindexofmultipledeprivationimdageofmothertypeofoutcomeandpreviousliverbornchildren.

21. Office for National Statistics. English indices of deprivation 2019 [Available from: https://www.gov.uk/government/collections/english-indices-of-deprivation.

22. Johnson RD, Ford DV, Broadhurst K, Cusworth L, Jones KH, Akbari A, *et al*. Data Resource: population level family justice administrative data with opportunities for data linkage. Int J Popul Data Sci. 2020;5(1):1339. https://doi.org/10.23889/ijpds.v5i1.1339

23. Wall-Wieler E, Roos LL, Bolton J, Brownell M, Nickel N, Chateau D. Maternal Mental Health after Custody Loss and Death of a Child: A Retrospective Cohort Study Using Linkable Administrative Data. Can J Psychiat. 2018;63(5):322–8. https://doi.org/10.1177/0706743717738494

24. Griffiths LJ, Johnson RD, Broadhurst K, John A. Born into care: One thousand mothers in care proceedings in Wales. London: Nuffield Family Justice Observatory; 2021. https://www.nuffieldfjo.org.uk/resource/1000-mothers-care-proceedings-wales.

25. Jay M, Mc Grath-Lone L, Pearson R, Gilbert R. The cumulative proportion of children receiving social care services in England: a whole population administrative data cohort study. International Journal of Population Data Science. 2022;7(3). https://doi.org/10.23889/ijpds.v7i3.1895

26. Mc Grath-Lone L, Dearden L, Nasim B, Harron K, Gilbert R. Changes in first entry to out-of-home care from 1992 to 2012 among children in England. Child Abuse Negl. 2016;51:163–71. https://doi.org/10.1016/j.chiabu.2015.10.020

27. Ubbesen MB, Gilbert R, Thoburn J. Cumulative incidence of entry into out-of-home care: changes over time in Denmark and England. Child Abuse Negl. 2015;42:63–71. https://doi.org/10.1016/j.chiabu.2014.10.006

28. Office for National Statistics. Conceptions in England and Wales: 2020. 2022. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/bulletins/conceptionstatistics/2020#teenage-conceptions.

29. Public Health England. Teenage mothers and young fathers: support framework. 2016. https://www.gov.uk/government/publications/teenage-mothers-and-young-fathers-support-framework.

30. Public Health England. A framework for supporting teenage mothers and young fathers 2019. https://assets.publishing.service.gov.uk/media/5cb85bc640f0b649e47f2983/PHE_Young_Parents_Support_Framework_April2019.pdf.

31. Barnardo's. Not the end of the story: Supporting teenage mothers back into education. 2010. https://www.barnardos.org.uk/sites/default/files/2020-11/not_the_end_of_the_story_-_march_2010_pdf.pdf.

32. Franklin J, Larkham J, Mansoor M. The well-worn path. 2023. https://www.childrenssociety.org.uk/information/professionals/resources/the-well-worn-path#facts-about-childrens-services-funding.

33. Knight HE, Gurol-Urganci I, van der Meulen JH, Mahmood TA, Richmond DH, Dougall A, Cromwell DA. Vaginal birth after caesarean section: a cohort study investigating factors associated with its uptake and success. BJOG. 2014;121(2):183–92. https://doi.org/10.1111/1471-0528.12508

34. Office for National Statistics. Births by parents' country of birth, England and Wales: 2021. 2022.https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/parentscountryofbirthenglandandwales/2021#:..:text=Main%20points,decrease%20from%2029.3%25%20in%202020.

## Abbreviations

| | |
|---|---|
| Cafcass: | Children and Family Court Advisory and Support Service |
| CI: | Confidence Interval |
| CP: | care proceedings |
| FLB: | First live birth |
| HES: | Hospital Episode Statistics |
| HES APC: | Hospital Episode Statistics Admitted Patient Care |
| ID: | Identifiers |
| IMD: | Index of multiple deprivation |
| LA: | Local Authority |
| MT: | Maternity Tail |
| NHSE: | NHS England |
| ONS: | Office for National Statistics |
| PDS: | Personal Demographic Service |
| UCL: | University College London |
| UCL DSH: | UCL Data Safe Haven |

# Supplementary Material

# 1. Data sources: Cafcass

## 1.1 Definitions relevant to family court data

| Term | Definition |
|---|---|
| Family proceedings | Family proceedings are cases heard by the family courts concerning the upbringing of children. In all cases, the paramount consideration is the welfare of the child.<br>In **public law** cases, also called **care proceedings**, the court determines applications brought by local authorities (through their children's social care services) for **care** or **supervision orders** (Children Act 1989, section 31). In public proceedings, the court can make a care or supervision order as well as any other order, such as the private law **section 8 orders**. An application for an order under section 31 can only be made by a local authority, who is responsible for completing the application form (current designation C110A). In public proceedings, the parents are entitled to non-means-tested legal aid and representation, as is the child.<br>In **private law** cases, the court determines applications made by private individuals concerning the upbringing of children (such as where a child will live following relationship breakdown or regarding specific issues). Private proceedings are predominantly for child arrangements **orders under section 8** of the Children Act 1989. In private proceedings, the court cannot make a care or supervision order. However, the court does have the power to ask the local authority to investigate and consider applying for a care order. Means-tested legal aid is only available in some circumstances such as domestic abuse; otherwise applicant and respondent are responsible conduct the litigation themselves or paying for legal representation.<br>In all proceedings, the court must consider whether making an order is better than no order at all; if not, the court must not make an order. |
| Children Act 1989 | The Children Act 1989 reformed and consolidated the law relating to the protection of children from harm and the promotion of their welfare. It is the core piece of legislation concerning family proceedings. Under section 1 of the Act, the welfare of the child is the paramount consideration in all proceedings. In addition to making rules about parental responsibility and other matters, the Act provides for local authority services for children in need of support, for investigation of child maltreatment and for court proceedings. |
| Case and application | A court case may concern one or more children, one or more application for one or more orders and one or more legal outcomes. One mother can be involved in several court cases, for example if a case is brought for a child born after the initial case. |
| Children's social care services | Children's social care refers to all forms of personal care for children and young people who need extra support.<br>This includes when the child's needs are such that parents are not able to meet them at home or if parents' needs prevent them from looking after a child. It can also be because the child is at risk of harm from someone at home or outside their family. |
| Cafcass - The Children and Family Court Advisory and Support Service | Cafcass were established by section 11 of the Criminal Justice and Court Services Act 2000 and represent the interests of children and young people in the family court and independently advise the family courts about what is safe for children and in their best interests.<br>Cafcass represent all children in public proceedings. In private proceedings, Cafcass carry out safeguarding checks and cease involvement in most cases (approximately 70%) before the first hearing. |
| Care order | A care order enables the local authority to look after a child in state care. Where a care order is granted, parental responsibility is shared between the parents and local authority, meaning the authority is empowered to make decisions about the child. Placements under a care order could be at home with the parents, with kin or non-kin fosterers or residential or secure settings. The threshold for making an order is that the child is suffering or at risk of significant harm where the harm is attributable to the care given by the parents, or where the child is beyond parental control. |
| Supervisions order | A supervision order places a child under local authority supervision. The threshold is the same for care orders. Under a supervision order, parental responsibility is NOT shared with the authority. |

| Term | Definition |
|---|---|
| Section 8 order | Section 8 orders include child arrangement orders (which replaced contact orders and residence orders), specific issue orders and prohibited steps orders. They are brought by relevant parties such as parents, though family members may be supported by a local authority in order to avoid care proceedings. |
| Parental responsibility | The bundle of rights, duties, powers and responsibilities that a parent has towards a child and their property. The Children Act 1989 determines where parental responsibility exists (e.g., both married parents (or those in a civil partnership) have parental responsibility for a child when the child is born. |
| Party to proceedings | Under the Family Procedure Rules, in both private and public proceedings, the child's parents and anyone else with parental responsibility for the child will generally be considered "parties to proceedings." Other individuals can apply to the court to be made party to proceedings. |
| Accommodation under section 20 | Under section 20 of the Children Act, the local authority may, if the parents do not object, accommodate a child. Such a child becomes looked after under state care. This happens without recourse to the courts and therefore these cases are not captured in Cafcass data or this study. Approximately half of all children starting to be looked after each year are looked after under section 20, though these children may also experience court proceedings at other times. |

### 1.2 Data criteria for inclusion for women with care proceedings

**Definition of section 31 public law family court care proceeding**

- Public law case: at least 1 section 31 application associated with case
- First section 31 application: between 1$^{st}$ April 2007 to 31$^{st}$ December 2021
- Case status: allocated or completed

**Definition of a mother:**

- female
- parent to child in public law family court care proceeding
- party to care proceedings

**Included in analysis:**

- 15–50 years old at first start of case

**Excluded from analysis:**

- <15 years or >50 years old at start of first case
- No date of birth recorded for the mother

## 2. Linkage of Cafcass and HES identifiers by NHS England

## 2.1 Identifiers used in each data source

Supplementary Table 1: Availability of personal identifiers in Cafcass, Personal Demographic Service (PDS) and Hospital Episode Statistics (HES)

| Linkage identifiers | Data sources | | |
|---|---|---|---|
| | Cafcass | NHS England | |
| | Care proceedings | PDS | HES |
| First name(s) | X | X | |
| Middle name(s) | | | |
| Surname(s) | X | X | |
| Date of Birth | X | X | X |
| Sex | X | X | X |
| Address of residence | X | X | |
| Postcode of residence | X | X | X |
| Anonymised Cafcass ID | X | | |
| NHS number | | X | X |
| Anonymised UCL-HESID | | | X |

## 2.2 Master patient service (MPS) matching algorithm at NHS England

### Alphanumeric trace

A deterministic matching algorithm where the minimum required fields for matching are family name, date of birth and sex. And this stage can only be run if family name is provided.

### Algorithmic trace

A probabilistic matching algorithm that is run only if no match was found in alphanumeric trace. The minimum required fields for matching are date of birth, sex and postcode. For each Cafcass query record, a set of PDS candidate records are identified by blocking, and all candidate records are scored.

#### Blocking

The algorithmic trace considers the following blocks:

- Family name, given name, date of birth
- Family name, sex, date of birth, postcode
- Given name, sex, date of birth, postcode
- Date of birth, postcode, sex

Date of birth, postcode and sex need to be an exact match, although historic values are accepted.

A maximum of 50 candidate records from the blocking steps are retained, prioritising those that matched in the highest number of blocks.

#### Scoring

Scores are based on the similarity of features from the query record. It is calculated from the average of similarity scores of given and family names, date of birth, postcode and sex. Where values are missing, they are not included in the calculation.

Scoring for each variable is out of 100 and further information on the scoring rules for date of birth, sex and postcode are available in the Person ID handbook, section 3.5. Jaro-Winkler algorithm is used to score family and given name.

#### Ranking

MPS ranks matching combinations using the average of similarity scores for all non-null identifiers, although similarity

Supplementary Table 2: Number of mothers with a section 31 care proceedings started between 1st April 2007 and 31st December 2021 in England who do not have a date of birth or postcode recorded

| | Mothers* | date of birth | | postcode | |
|---|---|---|---|---|---|
| | | | | n | % |
| *Total* | 120,937 | 5,935 | 4.9 | 10,882 | 9.0 |
| *Ethnicity* | | | | | |
| White | 73,889 | 1,974 | 2.7 | 5,506 | 7.5 |
| Black | 5,104 | 414 | 8.1 | 700 | 13.7 |
| Asian | 3,465 | 240 | 6.9 | 432 | 12.5 |
| Mixed | 3,245 | 113 | 3.5 | 271 | 8.4 |
| Other | 1,162 | 124 | 10.7 | 219 | 18.8 |
| Missing | 34,072 | 3,070 | 9.0 | 3,754 | 11.0 |
| *Year of first s31 case* | | | | | |
| 2007-2010 | 23,540 | 2,309 | 9.8 | 2,125 | 9.0 |
| 2011-2014 | 32,049 | 1,691 | 5.3 | 2,425 | 7.6 |
| 2015-2018 | 39,493 | 1,315 | 3.3 | 3,441 | 8.7 |
| 2019-2021 | 25,855 | 620 | 2.4 | 2,891 | 11.2 |
| *Number of cases* | | | | | |
| 1 | 94,405 | 5,668 | 6.0 | 10,266 | 10.9 |
| 2 | 19,490 | 244 | 1.3 | 566 | 2.9 |
| 3+ | 7,042 | 23 | 0.3 | 50 | 0.7 |
| *Region of first case*& | | | | | |
| North East | 9,855 | 395 | 4.0 | 1,100 | 11.2 |
| North West | 19,840 | 555 | 2.8 | 1,730 | 8.7 |
| East Midlands | 9,415 | 320 | 3.4 | 725 | 7.7 |
| Yorkshire and Humber | 13,915 | 575 | 4.1 | 830 | 6.0 |
| West Midlands | 13,010 | 680 | 5.2 | 1,230 | 9.5 |
| East of England | 10,570 | 715 | 6.8 | 820 | 7.8 |
| London | 17,625 | 1,625 | 9.2 | 2,350 | 13.3 |
| South East | 15,755 | 665 | 4.2 | 1,285 | 8.2 |
| South West | 10,410 | 375 | 3.6 | 760 | 7.3 |
| Missing | 545 | 35 | 6.4 | 55 | 10.1 |

Supplementary Table 3: Unadjusted and adjusted odds ratio of linking to the Delivery Cohort for mothers party to care proceedings between 1st April 2007 and 31st December 2021, and available for linkage (n = 112,147), in England

| | Unadjusted odds ratio | | Adjusted odds ratio (n = 112,147) | |
|---|---|---|---|---|
| | **OR** | **95% CI** | **OR** | **95% CI** |
| *Ethnicity* | | | | |
| White | **Ref** | | **Ref** | |
| Black | 0.38 | 0.36–0.41 | 0.54 | 0.50–0.58 |
| Asian | 0.51 | 0.47–0.55 | 0.64 | 0.58–0.69 |
| Mixed | 0.68 | 0.62–0.74 | 0.74 | 0.67–0.81 |
| Other | 0.31 | 0.28–0.36 | 0.43 | 0.37–0.49 |
| Missing | 0.67 | 0.65–0.69 | 0.83 | 0.79–0.86 |
| *Number of children* | | | | |
| 1 | **Ref** | | **Ref** | |
| 2 | 1.36 | 1.31–1.42 | 1.24 | 1.19–1.29 |
| 3 | 1.57 | 1.50–1.65 | 1.39 | 1.32–1.46 |
| 4+ | 1.76 | 1.67–1.85 | 1.49 | 1.41–1.58 |
| Age at first s31 case | | | | |
| 15–19 | 0.92 | 0.86–0.98 | 0.99 | 0.93–1.05 |
| 20–24 | **Ref** | | **Ref** | |
| 25–29 | 0.94 | 0.89–0.99 | 0.92 | 0.87–0.97 |
| 30–34 | 0.83 | 0.79–0.88 | 0.82 | 0.78–0.87 |
| 35–39 | 0.68 | 0.64–0.72 | 0.71 | 0.67–0.75 |
| 40–44 | 0.50 | 0.47–0.53 | 0.56 | 0.53–0.59 |
| 45–50 | 0.37 | 0.35–0.4 | 0.43 | 0.40–0.46 |
| *Year of first s31 case* | | | | |
| 2007–2010 | 0.90 | 0.86–0.95 | 0.85 | 0.81–0.89 |
| 2011–2014 | **Ref** | | **Ref** | |
| 2015–2018 | 1.03 | 0.99–1.07 | 1.06 | 1.01–1.11 |
| 2019–2022 | 1.14 | 1.09–1.20 | 1.24 | 1.18–1.31 |
| Region of first case | | | | |
| North East | 1.37 | 1.27–1.47 | 1.30 | 1.20–1.40 |
| North West | 1.04 | 0.98–1.10 | 1.01 | 0.96–1.08 |
| Yorkshire and the Humber | 1.26 | 1.18–1.35 | 1.19 | 1.11–1.27 |
| East Midlands | 1.15 | 1.07–1.24 | 1.11 | 1.03–1.20 |
| West Midlands | 1.06 | 0.99–1.13 | 1.05 | 0.99–1.13 |
| East of England | 0.86 | 0.81–0.92 | 0.85 | 0.79–0.91 |
| London | 0.55 | 0.52–0.59 | 0.70 | 0.66–0.74 |
| South East | **Ref** | **Ref** | | |
| South West | 1.04 | 0.97–1.11 | 1.01 | 0.94–1.08 |
| Not available | 0.82 | 0.66–1.03 | 0.81 | 0.64–1.02 |
| Number of s31 cases | | | | |
| 1 | **Ref** | | **Ref** | |
| 2 | 1.53 | 1.46–1.60 | 1.18 | 1.12–1.24 |
| 3+ | 1.76 | 1.64–1.90 | 1.18 | 1.08–1.28 |
| Postcode available | | | | |
| Yes | 1.98 | 1.89–2.08 | 1.77 | 1.68–1.86 |

scores need to be >50 for consideration. The highest score is taken forward. However, if 2+ match-combinations have the highest ranking scores, or similar score (within 5), no match is returned.

## Returned to UCL in linkage bridging file

NHSE returned a linkage bridging file to UCL-DSH containing anonymised Cafcass and UCL-HES IDs only. The file contained

UCL-HES IDs only where a Cafcass name and address combination matched uniquely to a UCL-HES ID. Each women in Cafcass could have up to 3 addresses in the Cafcass Linkage File, a line for each record, and could therefore match to up to 3 UCL-HES records.

Within UCL Data Safe Haven, Cafcass-HES match combinations were reviewed.

## 2.3 Odds of matching

Supplementary Table 2 shows the characteristics of women with a section 31 care proceeding between 1st April 2007 and 31st December 2021 who did not have a date of birth, and were excluded from linkage as a result, and who had no postcode recorded.

Supplementary Table 3 shows the unadjusted and adjusted odds of linking to a delivery record in UCL-HES extract for mothers involved in care proceedings (92,891/112,147).

# 3. Data sources: hospital episode statistics to derive a delivery cohort -extended methodology

Here we describe in detail the process through which the delivery cohort (DC) was created. The final cohort includes women aged between 15 and 50 at delivery with a gestation $\geq$24 weeks (or not reported) who deliver in English NHS units between 1$^{st}$ April 1997 and 31$^{st}$ December 2022 and are resident in England. We describe the four stages of cohort generation:

1) the extraction of all deliveries from HES APC
2) the process through which sequential deliveries for a woman are identified and dates of delivery are estimated
3) how delivery characteristics are coded, and,
4) criteria used to exclude women from the cohort.

There are additional ways in which to improve the quality of data with the delivery cohort, for example by linking delivery and baby records, as described by Harron et al. (2016) [19], however, this is not explored within this manuscript.

## 3.1 Stage 1: Extracting delivery episodes from HES APC

Inpatient admissions in HES APC are made up of 1 or more consultant 'episodes', defined as the time during which a patient it under the care of one hospital consultant. Within HES APC, episodes related to the delivery of an infant(s) (live or stillborn) can be identified, and maternal demographic and clinical details can be extracted. Additional information is provided as part of a 'maternity tail' (MT), which includes information on the mother, such as gestational age, number of previous pregnancies, mode of delivery, and up to 9 infants, such as sex, status at delivery and birth weight. Clinical coders use medical discharge notes to code diagnostic codes (using International Classification of Diseases, tenth revision [ICD-10]), proceedural codes (using Office of Population Censuses and Surveys Classification of Interventions and Procedures version 4 [OPCS-4]) and MT variables (listed in Supplementary Table 4).

To create this cohort, all available delivery episodes for women aged between 12 and 50 with a gestation $\geq$24 weeks (or not recorded) in NHS provider hospitals in England were extracted from all available years of HES APC at University College London, using the criteria in Supplementary Table 5.

## 3.2 Stage 2: identifying sequential deliveries and deriving estimated date of delivery

Infant date of birth is not available in anonymised HES APC datasets, as a result, we devised the below methodology to identify episodes associated with a delivery and the episode amongst those that was most likely to contain the delivery. Using this inferred delivery episode, we could estimate delivery date.

### Delivery 1

The process to estimate delivery date for delivery 1 in HES APC is described in more detail below and example episode timeline shown in Supplementary Figure 1.

1. *Identify first delivery episode and generate a 'delivery window' to identify and process all episodes that could be associated with delivery 1*

The first delivery episode for each woman was identified within the data, and a 'window' of 150 days from this date used to identify all episodes that could be associated with pregnancy 1 within the data. As delivery episodes were extracted from where 24 weeks gestation was recorded in gestat_1, 150 days was used to include women who had multiple admissions over their pregnancy but may have eventually delivered closer to or at full term. The majority (98.4%) of women had only 1 episode within the delivery window. Using a sample of 1/5$^{th}$ of first deliveries to compare the first and last episode date within the delivery window, 97.2% were less than 100 days apart and 99.2% were less than 130 days.

2. *Within delivery window, identify episode most likely to include delivery of infant*

If a women had only 1 episode within the delivery window, this was marked as their delivery episode.

For women with more than one episode within their delivery window (1.6% of deliveries), the following processing steps aimed to find the episode that was most likely to have contained the delivery, judged by the episode with the most information on the infant and delivery.

Firstly, for every episode in the delivery window the number of MT variables that contained information (not null) were counted to provide a MT completion score (Supplementary Figure 2a). The highest MT completion score per women for the delivery window was identified and episodes with lower scores were discarded, as it was assumed that an episode with more MT information available would be more likely to have contained the delivery itself. If this left 1 episode in the delivery window, it was marked as the delivery episode.

Supplementary Table 4: Maternity Tail (MT) variables

| Variable name | Definition |
|---|---|
| *anagest* | gestation period in weeks at first antenatal assessment |
| *anasdate* | first antenatal assessment date |
| *antedur* | antenatal days of stay |
| *biresus_1 to biresus_9* | resuscitation method for baby 1 to 9 |
| *birordr_1 to birordr_9* | birth order for baby 1 to 9 |
| *birstat_1 to birstat_9* | birth status for baby 1 to 9 |
| *birweit_1 to birweit_9* | birth weight for baby 1 to 9 |
| *delchang* | delivery place change reason |
| *delinten* | intended delivery place |
| *delmeth_1 to delmeth_9* | alternative delivery method for baby 1 to 9- derived |
| *delonset* | labour/delivery onset method |
| *delposan* | anaesthetic given post-labour or delivery |
| *delprean* | anaesthetic given during labour or delivery |
| *delstat_1 to delstat_9* | status of person conducting delivery for baby 1 to 9 |
| *gestat_1 to gestat_9* | length of gestation for baby 1 to 9 |
| *matage* | mothers age at delivery |
| *neocare* | Level of care given to a newborn child |
| *numbaby* | Number of babies delivered at end of single pregnancy |
| *numpreg* | Number of previous pregnancies resulting in a registrable birth |
| *postdur* | postnatal stay |
| *sexbaby_1 to sexbaby_9* | sex of baby 1 to 9 |

Where more than 1 episode remained within the 'delivery window', antedur (a derived variable containing the number of days between episode start and infant date of birth) was used to identify episodes closest to delivery date for each woman (Supplementary Figure 2b). Remaining episodes within the 'delivery window' without the lowest antedur value for each woman, including if they were null, were discarded. If this left 1 episode, it was marked as the delivery episode.

Where more than 1 episode remained within the 'delivery window', postdur (a derived variable containing the number of days between episode end date and infant date of birth) was used to identify episodes closest to delivery date (Supplementary Figure 2c). Remaining episodes within the 'delivery window' without the lowest value for each woman, including if they were null, were discarded. If this left 1 episode, it was marked as the delivery episode.

Where more than 1 episode remained within the 'delivery window', the number of delivery procedures coded (R17-25 & R27, therefore excluding induction procedures) for each episode were counted and the highest count per woman identified. Episodes within the 'delivery window' without the highest count for each woman were discarded (Supplementary Figure 2d). If this left 1 episode, it was marked as the delivery episode.

Following this process, 0.2% of women still had more than one episode remaining within the delivery window. In these cases, the first episode with a live infant recorded, or first episode within the delivery window was taken.

3. *Estimate delivery date*

Once the episode most likely to contain the delivery is identified for each woman, the date of delivery was estimated using the following steps:

i. Where antedur (a derived variable containing the number of days between episode start and infant date of birth) available, episode start date + antedur value, else

ii. Where postdur (a derived variable containing the number of days between episode end date and infant date of birth) available, episode end date - postdur value, else

iii. where delivery procedure available, use first delivery procedure date within episode, else

iv. Episode start date.

**Subsequent deliveries (2+)**

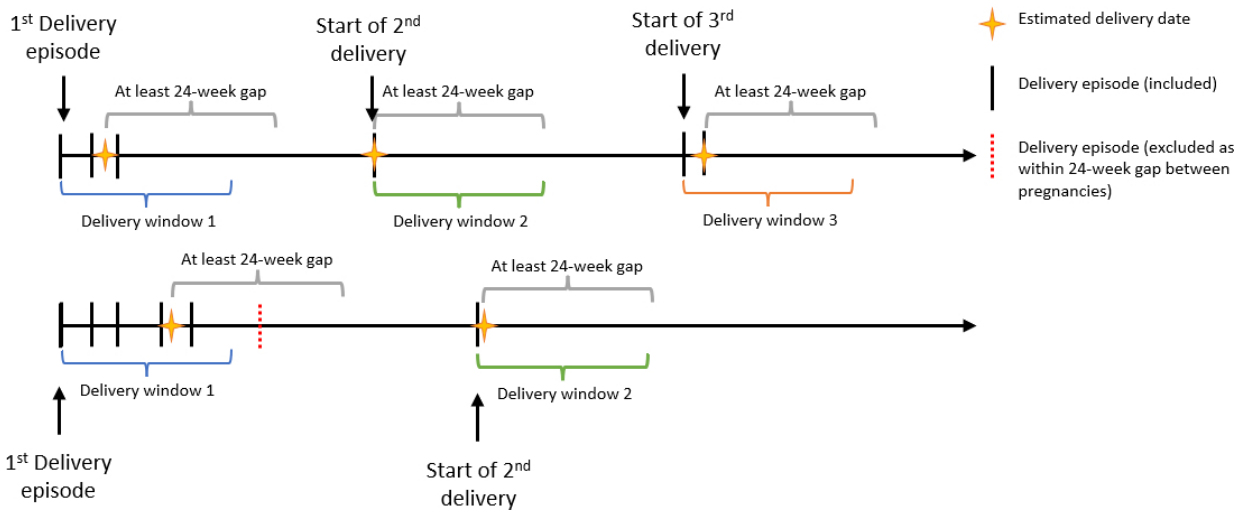
Once a date of delivery was estimated for delivery 1, the first episode associated with a second delivery could be identified, where available, for each woman. The subsequent delivery had to be at least 24 weeks after the first delivery date, and if episodes occurred before this time they were excluded.

As with delivery 1, a 'delivery window' was created to capture any episodes within a 150 days of the of the first (first episode start date + 150 days). Following which, the process to identify the episode most likely to include the delivery, estimate delivery date and identify subsequent deliveries was repeated until no episodes remained.

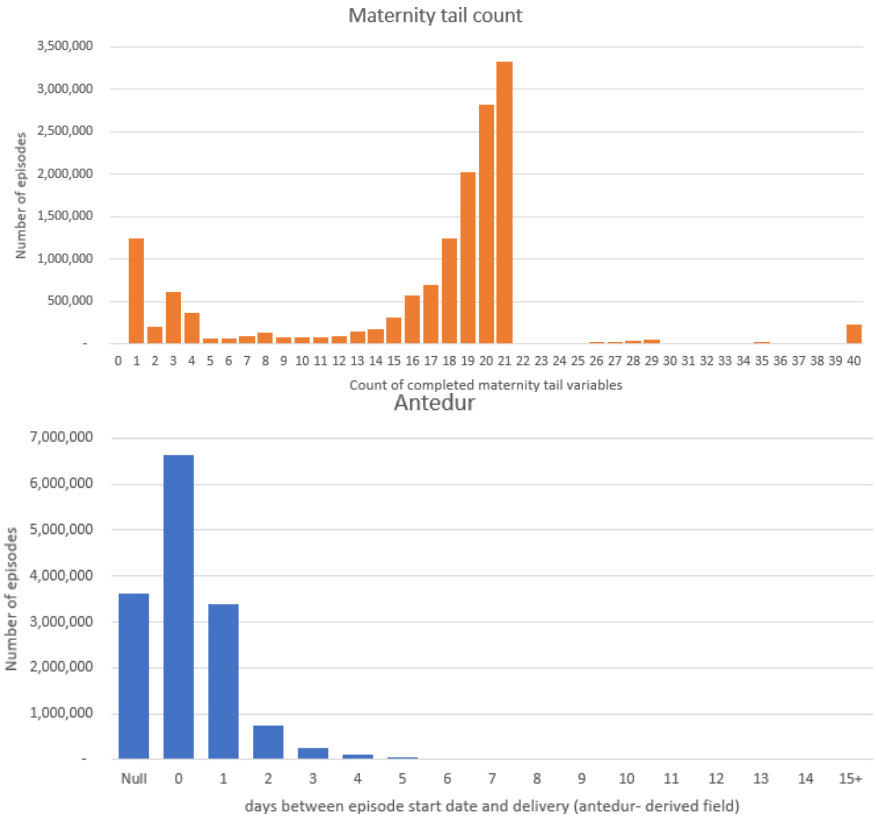
Supplementary Table 5: Criteria for identifying delivery episodes within Hospital Episode Statistics Admitted Patient Care data

| **Variable** | **Variable name in HES APC** | **Criteria** | |
|---|---|---|---|
| Proceedural codes (OPCS-4.10) | *optern_1->optern_** | R14 – surgical induction of labour<br>R15 – other induction of labour<br>R17 – Elective caesarean delivery<br>R18 – Other caesarean delivery<br>R19 – breech extraction delivery<br>R20 – other breech delivery<br>R21 – forceps cephalic delivery<br>R22 – vacuum delivery<br>R23 – cephalic vaginal delivery with abnormal presentations of head at delivery without instruments<br>R24 – normal delivery<br>R25 – other methods of delivery<br>R27 – other operations to facilitate delivery | OR |
| ICD- 10 diagnostic code | d*iag_1->diag_** | Z37 – outcome of delivery<br>Z38 – liveborn infants according to place of delivery | OR |
| Delivery method | *delmeth_1* | 0 = Spontaneous vertex (normal vaginal delivery, occipitoanterior)<br>1 = Spontaneous other cephalic (cephalic vaginal delivery with abnormal presentation of head at delivery, without instruments, with or without manipulation)<br>2 = Low forceps, not breech, including forceps delivery not otherwise specified (forceps, low application, without manipulation)<br>3 = Other forceps, not breech, including high forceps and mid forceps (forceps with manipulation)<br>4 = Ventouse, vacuum extraction<br>5 = Breech, including newlinetial breech extraction (spontaneous delivery assisted or unspecified)<br>6 = Breech<br>7 = Elective caesarean section<br>8 = Emergency caesarean section<br>9 = Other | AND |
| Sex | *sex* | 2 (female) | AND |
| Age at start of episode | *startage* | Between 12 and 50 | AND |
| Episode end date | *epiend* | Is not null | AND |
| Episode status | *epistat* | 3 = finished | AND |
| Provider code | *procode* | Does not begin with 'N' or '8'<br>N* = typically privately run providers<br>89997 = UK provider where no organisation code has been issued<br>89999 = Non-NHS UK provider where no organisation code has been issued | AND |
| Length of gestation | *gestat_1* | Is null or ≥24 | AND |
| At least 2 of the following maternity tail variables are completed | *anagest, anasdata, antedur, biresus_1, birordr_1, birstat_1, birweit_1, delchang, delinten,delmeth_1, delonset, delposan, delprean, delstat_1, gestat_1, matage, postdur, sexbaby_1* | At least two of these are not null | |

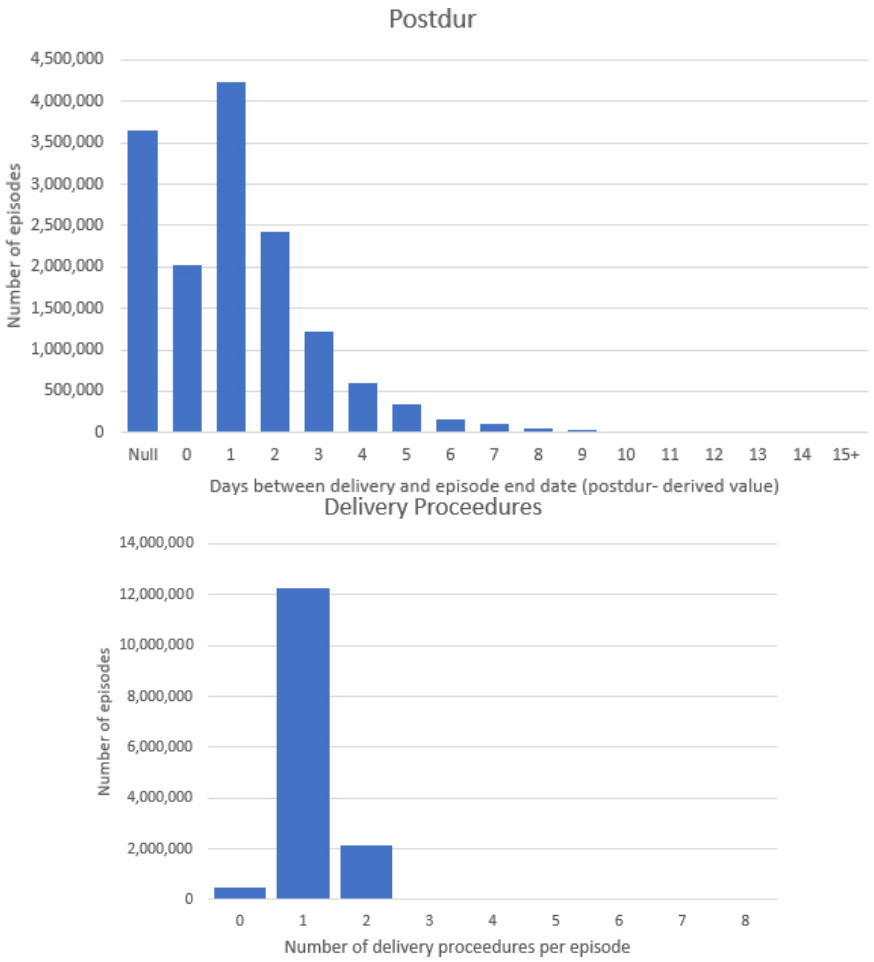
Supplementary Figure 1: An example timeline of how delivery episodes and estimated delivery date may look over a women's delivery episode history




Supplementary Figure 2: Within all extracted episodes, frequency histograms of a) count of maternity tail variables that contain information (MT completion score) for each episode, b) the number of days between the start of the episode and infant date of birth (antedur values), c) the number of days between infant date of birth and end of episode (postdur values) and d) count of delivery procedures recorded for each episode



Supplementary Figure 2: Continued




## 3.3 Stage 3: Coding delivery characteristics and exclusions

### Coding number of infants delivered, birth status and previous pregnancies

Data recorded in the diagnostic coding and MT variables can be used to describe delivery characteristics. For the majority, there is only one variable through which to ascertain the information, such as birth weight or gestation, but for the number of infants delivered, birth status and number of previous pregnancies there is more than one potential source of information.

#### Birth multiplicity/Number of infants delivered

Diagnostic codes and MT variables can be used to code the number of infants delivered (Supplementary Table 6), however variables conflicted in 26% of deliveries.

Where there were conflicts across the variables, the patterns of variable combinations were assessed, sometimes in conjunction with the provider code. We made decisions on how to code them after reviewing a sample of the records, examining consistency and comparing it to national estimates of the proportion of women who delivered multiples. Diagnosis codes were given more weight over maternity tail variables, particularly *numbaby*, because the review highlighted issues along the maternity tail, which include:

- a disproportionate number of women recorded as delivering between 4 and 9 infants, the results of the same information being copied across MT variables_1-9 and numbaby being a derived variable based on counting the number of variables with data in (usually *birordr_1 to birordr_9*) ;
- when '*numbaby*' corresponded with diagnosis codes, the amount of infant specific information across other variables, such as '*birstat*' or '*sexbaby*', was frequently lower than expected.

Exceptions to when diagnosis codes were provided with more weight, were limited to records where more birthweights of different values were recorded than diagnosis codes suggested.

There is opportunity to improve the coding of number of infants delivered through the Mother-Baby linkage (Harron et al. 2016 [19]), but this is not included within this manuscript.

#### Birth status

The *birstat_1-9* and *diag_01-24* variables were used to ascertain whether deliveries contained live or still born infants (Supplementary Table 7). Where *birstat* and *diag* codes

Supplementary Table 6: Maternity tail and diagnosis codes used to code the number of infants delivered

| Variable | Singleton | Twins | Other multiple |
|---|---|---|---|
| Birstat_*/sexbaby_*/ Birorder_*/ birweit_* | _1 is not null and _2 is null | _2 is not null and _3 is null | _3+ is not null |
| Numbaby | 1 | 2 | $\geq 3$ |
| Diag_* | Z37.0=single live birth<br>Z37.1 = single still birth<br>Z38.0=singleton, born in hospital<br>Z38.1=singleton, born outside hospital<br>Z38.2=singleton, not specified | Z37.2 = twins, both liveborn<br>Z37.3 = twins, one liveborn one stillborn<br>Z37.4 = twins, both stillborn<br>Z38.3=twin, born in hospital<br>Z38.4=twin, born outside hospital<br>Z38.5=twin, unspecified | Z37.5 = other multiple births, all liveborn<br>Z37.6 = other multiple births, some liveborn<br>Z37.7 = other multiple births, all stillborn<br>Z38.6=other multiple, born in hospital<br>Z38.7=other multiple, born outside hospital<br>Z38.8=other multiple, unspecified |

suggested conflicting outcomes (0.24%; n = 35,338 deliveries), records were temporary coded as 'uncertain' and reviewed.

During the review process, the proportion of deliveries coded as live, still and uncertain for a provider by year were tabulated. Where providers had more than 3% of deliveries coded as uncertain in a given year, quality of coding was deemed to be low. These records typically had a diagnosis codes suggesting a live birth and *birstat* suggesting a still birth. For the records within these providers and specific years, they were coded a live births. All other records remained coded as uncertain.

There is opportunity to improve the coding of infant birth status through the Mother-Baby linkage (Harron et al. 2016 [19]), but this is not explored within this manuscript.

### Number of previous pregnancies

The sequential processing of deliveries provides information on the number of previous pregnancies within HES APC records; however, the data is biased by left censoring. Most significantly, the result of deliveries prior to data collection in 1997. Alternate scenarios include; when women migrate to England with children or previous pregnancies and then deliver in NHS hospitals, women who previously delivered within private hospitals and women who delivered in NHS hospitals in England and were resident at the time who didn't link within HES.

Where available, *numpreg*, the number of previous pregnancies resulting in a registrable birth (live or still born), was used to update delivery order. *Numpreg* was available for 60.5% of deliveries, and somewhere on the record for 68.1% of women. To recalibrate the number of previous pregnancies, the first delivery for each woman with information in the variable *numpreg* was used. This was available for 5.64 million women, 80% of whom it was available on their first delivery within the cohort.

It is estimated that between 40-45% of deliveries in England each year are in primiparous women. Prior to updating the number of previous pregnancies with *numpreg*, the frequency of *numpreg* values for each provider per year was assessed for data quality. This highlighted a small number of providers and years where the only returned *numpreg* values were between 10–15%, the information from which were excluded (n = 2,384 women).

Additionally, where less than 10% of *numpreg* values were 0 for provider/years but the number with 1 previous pregnancy

Supplementary Table 7: Birth Status and ICD-10 codes used to ascertain live born status

| Variable | Live born | Live birth and still birth | Still born |
|---|---|---|---|
| *birstat_1–9* | 1 = live birth | | 2 = still birth (antepartum)<br>3 = still birth (intrapartum)<br>4 = still birth (indeterminate) |
| *diag_01–20* | Z37.0 = single live birth<br>Z37.2 = twins, both liveborn<br>Z37.5 = other multiple births, all liveborn | Z37.3 = twins, one liveborn one stillborn<br>Z37.6 = other multiple births, some liveborn | Z37.1 = single still birth<br>Z37.4 = twins, both stillborn<br>Z37.7 = other multiple births, all stillborn<br>O00, O03-O08 or P964 & gestat_1 is $\geq$24 weeks |

was high, 1 was subtracted from the recorded *numpreg* value (n = 368,586 women). This pattern suggests some providers were miscoding women with no previous pregnancies (coded as 1 but should be 0). Amongst the women with a *numpreg* value, 2.1% had a *numpreg* value lower than the processed HES data delivery figure and number of previous pregnancies were not revised.

Following this, *numpreg* was available for 66.1% of mothers and delivery sequences were recalibrate (up only) the number of previous pregnancies for 20.4% of women. In 45.7% of mothers, numpreg was the same as delivery sequence. Prior to adjustment, 57% of deliveries would have been classified as a first birth- which was reduced to 45%.

This methodology does not allow adjustment further down the delivery sequence where a middle child was born outside of HES APC but a subsequent delivery is included in the cohort.

### 3.4 Stage 4: Excluded records

Mothers were excluded from the cohort for a number of reasons. As maintaining confidence in the sequence of deliveries for each mother in the cohort was important, if one delivery record met one of these criteria their whole delivery history was excluded from the cohort. The reasons for exclusion were:

- **Under 15 years at first delivery in cohort**: Whilst the delivery records of 12-14 year olds were processed, women who delivered under 15 years were excluded due to concerns of deductive disclosure when analysed in conjunction with data from care proceedings.
- **Delivery episodes less than 24 weeks after an estimated delivery date**: Following processing, there were 2,001 women with delivery episodes recorded less than 24 weeks after an estimated delivery date. These women were excluded from the cohort.
- **Data quality in specific providers and years**: Upon review of delivery records that occur within 24 weeks of an estimated delivery date, we realised there are some issues with coding between 1997 and 2001 in specific providers. These providers seem to have a high number of delivery episodes where delmeth_1 is available within the maternity tail, but the episode has pregnancy ICD-10 codes, rather than delivery diagnoses or procedure codes. The trend in delivery episodes in these providers also suggested a sudden fall in deliveries per year at a specific time point. Following this, delivery numbers by year were reviewed and women with a delivery in 15 providers between 1997 and 2001 (years excluded were specific to the pattern of coding for the provider) were excluded from the cohort. Amongst women in these providers and years, 20.3% of final delivery episodes had no delivery procedure codes on the record, compared to 2.2% of all other delivery records in the cohort.
- **Delivery records with abortion codes**: Delivery episodes were removed where they had abortion ICD-10 codes (O01, O03-08, P964) listed but no gestation recorded. Where gestation was recorded as $\geq$24 weeks, they were retained in the cohort but recoded as still born.
- **Place of residence outside of England**: Women were excluded if their place of residence was recorded as outside of England.
- **Deliveries prior to 1st April 1997**: HES APC started from 1st April 1997, but some estimated delivery dates are prior to this date. Due to uncertain case ascertainment or data quality during this time women with deliveries prior to 1st April 1997 were excluded.
- **Large difference between episode start and delivery date**: Comparison of episode start year and estimated date of delivery show a small number of deliveries had a significant gap between them. This is likely the result of entry data errors in dates in HES APC. For 99.6% of deliveries, the difference was $\pm$10 days and 99.99% were $\pm$100 days. Women with deliveries where the difference was greater than 100 days were excluded.
- **Death was recorded prior to or between deliveries.**

Overall, 168,608 women had at least one delivery that met one of these criteria, resulting in 282,654 deliveries being excluded from the cohort.

Following linkage to Cafcass care proceedings an additional 455 women were excluded from the delivery cohort because their care proceedings occurred more than 39 weeks prior to a first live birth, or where they died before first care proceedings (n = 29).

### First live birth cohort

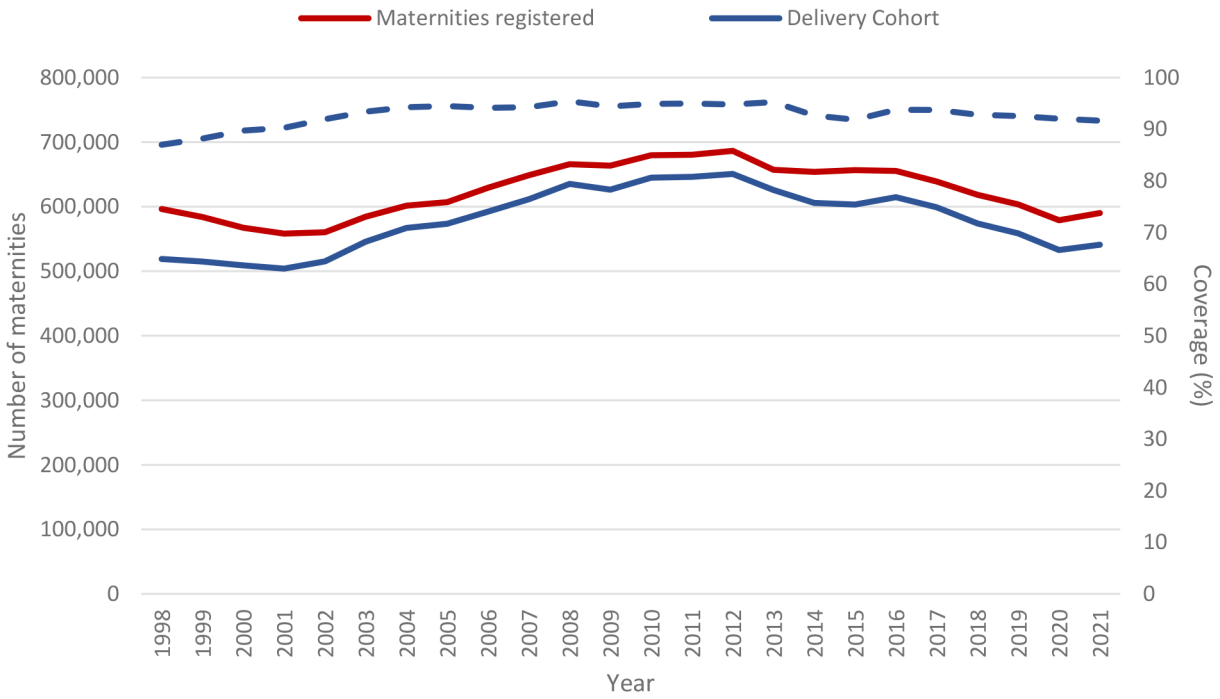
Women were included in the first live birth cohort if their first delivery record in the delivery cohort met the following criteria:

- Between 1st April 2007 and 31st December 2021
- Recorded as live birth
- Where quality of MT variable recording number of previous registrable pregnancies (live or still born) (aka numpreg) was high for a provider and its value suggest no previous pregnancies.
    - Refer to section 3.3 on how numpreg was used to adjust estimated number previous deliveries.

## 4. Comparison of delivery cohort to registrable maternities and births data recorded by the Office for National Statistics in England

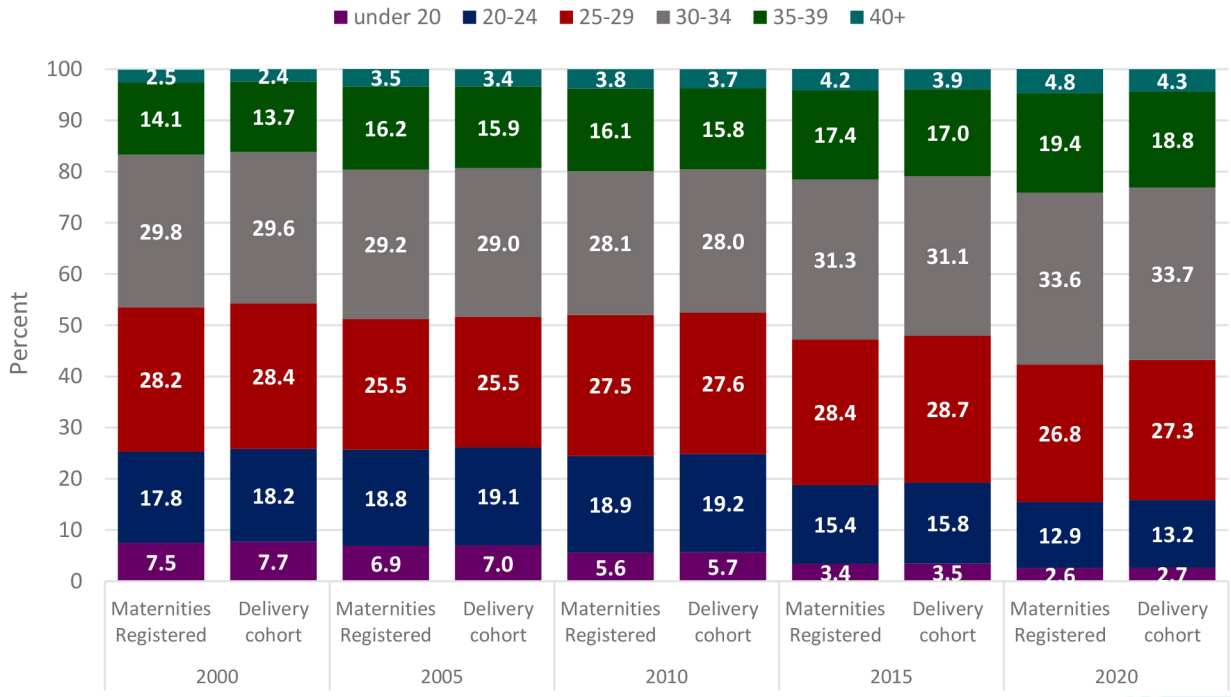
Demographic and delivery outcomes of mothers within the cohort were compared to civil registration of birth data for England, availabile from the Office for National Statistics (ONS), to determine whether the cohort was representative of all births in England (Supplementary Figures 3–7). To estimate coverage we compared the number of deliveries per year in the delivery cohort to those registered in England. We also compared the age and deprivation profile of mothers who delivered in England to the delivery cohort, and the proportion of all deliveries that contained more than 1 infant,

Supplementary Figure 3: Number of deliveries* in England by year between 1998 and 2021 registered (red line, source: ONS) and within the HES Delivery Cohort (blue line) and proportion of deliveries in England contained within the HES Delivery Cohort (coverage, dashed blue line, secondary y-axis)



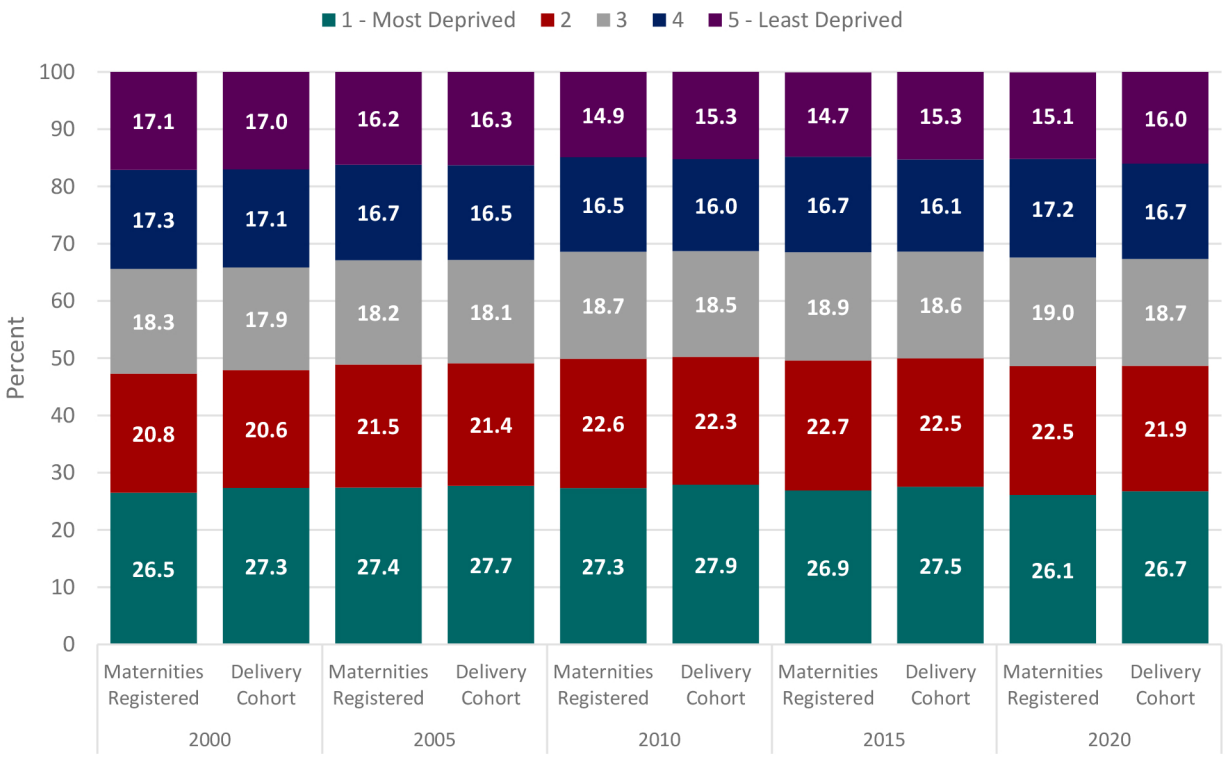

*delivery of live or still-born babies. If multiples were delivered only counted once.

Supplementary Figure 4: Comparison of age at delivery* in England by year for maternity registration data and Hospital Episode Statistics derived Delivery Cohort



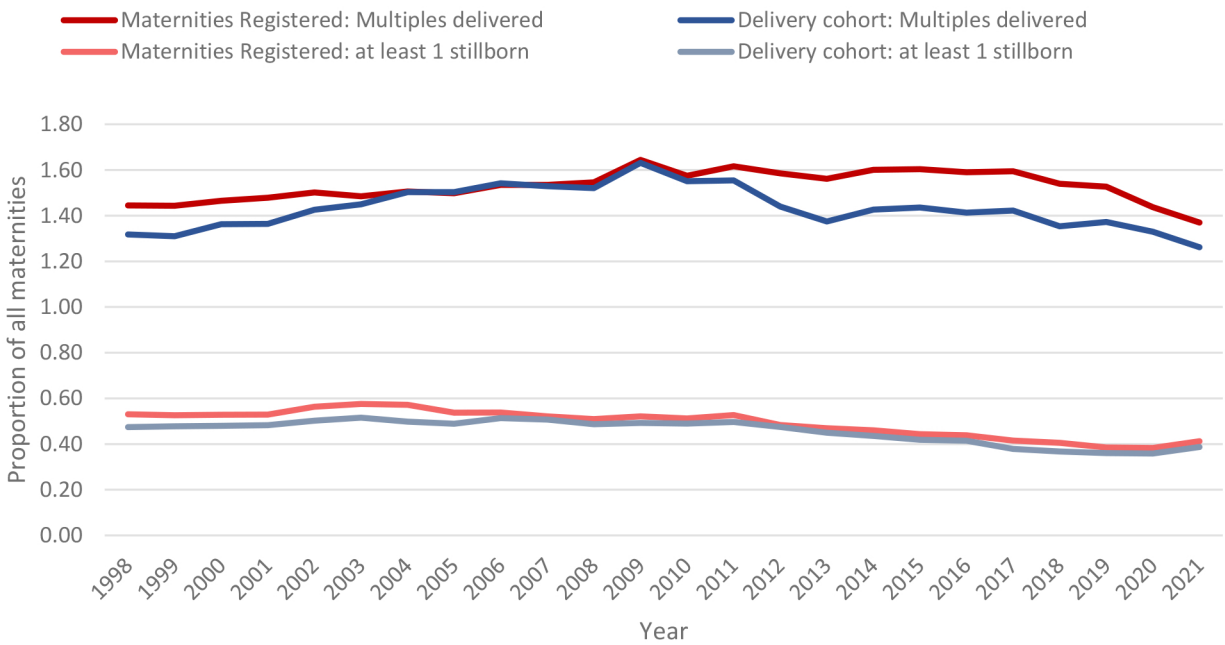

*delivery of live or still-born babies. If multiples were delivered only counted once.

Supplementary Figure 5: Comparison of deprivation quintile at delivery* in England by year for maternity registration data and Hospital Episode Statistics derived Delivery Cohort



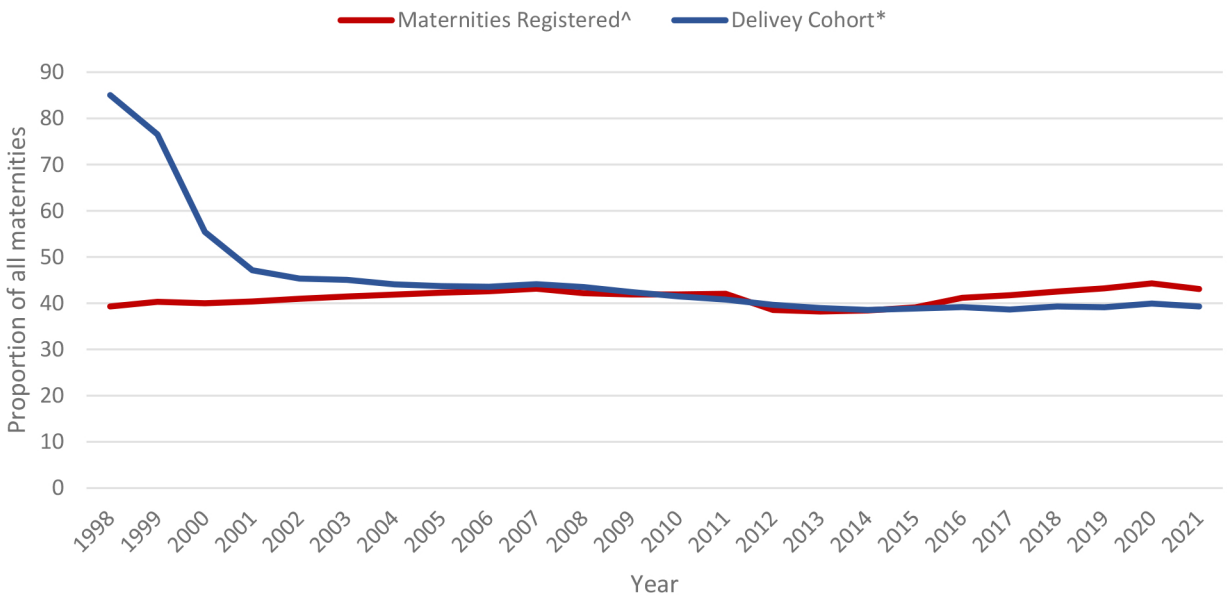

*delivery of live or still-born babies. If multiples were delivered only counted once.

Supplementary Figure 6: Comparison of the proportion of all deliveries* that contain more than 1 infant and at least 1 still born infant in England by year for maternity registration data (dark red and pink respectively) and Hospital Episode Statistics derived Delivery Cohort (dark and light blue lines)




*delivery of live or still-born babies. If multiples were delivered only counted once.

Supplementary Figure 7: Comparison of the proportion of women primiparous at delivery in England by year for maternity registration data and Hospital Episode Statistics derived Delivery Cohort




^previous live births.

*Using previous delivery admissions in HES and variable that, where recorded, states number of previous pregnancies that resulted in a registrable (live or stillborn) infant (HES APC variable name: numpreg).

at least 1 stillborn infant or were a first delivery (primiparous). Whilst data from 1st April 1997 to 31st December 2021 were processed to generate the delivery cohort, to mirror available data, we present data for the calendar years 1998-2021.

**Source:** Office for National Statistics (2023) Number of maternities by Index of Multiple Deprivation (IMD), age of mother, type of outcome and previous live born children. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/maternities/adhocs/1637numberofmaternitiesbyindexofmultipledeprivationimdageofmothertypeofoutcomeandpreviousliverbornchildren.

## 5. Evaluation of linkage quality: disagreement between common variables in HES and Cafcass

To provide indirect evidence of potential false links, we explored disagreement between characteristics that were recorded in both HES and Cafcass among mothers with a first live birth from 1st April 2007 who linked to Cafcass (Supplementary Table 8).

Supplementary Table 8: Disagreement on the values for six variables in linked healthcare and Cafcass records in The First Live Birth (FLB) Cohort (a), and five most common patterns of disagreement (b). The six variables compared were ethnicity, age at first delivery, number of live children at case start[1], region[2] and year of birth[3]. Score of 1 was given if the Cafcass and FLB values for a variable disagreed, 0 was given if the value was the same or one or both values were missing

**(a) Disagreement score between Cafcass and delivery records**

| Disagreement score | Number | Percent | Percent[4] with missing information for 0, 1 or 2 variables | | |
|---|---|---|---|---|---|
| | | | 0-no missing information | 1 | 2 |
| 0 – disagree on no variables | 28,694 | 74.6 | 75% | 25% | 0.1 |
| ***Disagree on 1 or more variables*** | ***9,768*** | ***25.4*** | 82% | 18% | <0.1 |
| 1 | 6,078 | 15.8 | 83% | 17% | <0.1 |
| 2 | 3,358 | 8.7 | 80% | 20% | <0.1 |
| 3 | 320 | 0.8 | 88% | 12% | |
| 4 | 12 | 0.0 | 100% | | |
| 5 – disagree on all variables | 0 | 0.0 | | | |
| Total | 38,462 | 100.0 | 77% | 23% | 0.1% |

**(b) Five most common patterns of disagreement**

| Ethnicity | Age at first delivery | Number of children at case[1] | Region[2] | Year of birth[3] | Disagreement Score | Frequency | Percent of disagreements[5] |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 0 | 1 | 3,337 | 34.2 |
| 0 | 1 | 1 | 0 | 0 | 2 | 2,725 | 27.9 |
| 1 | 0 | 0 | 0 | 0 | 1 | 1,224 | 12.5 |
| 0 | 1 | 0 | 0 | 0 | 1 | 753 | 7.7 |
| 0 | 0 | 0 | 1 | 0 | 1 | 733 | 7.5 |

1-based on number of children associated with a case and number of live children delivered within FLB Cohort prior to case start date. Where women have more than 1 case, they needed to disagree for all cases.
2-region was compared for all cases and deliveries and needed to disagree on all combinations.
3-not available within Cafcass deidentified data so estimated based on age at case start and date of case start. To disagree, FLB Cohort year of birth and Cafcass estimated year of birth had to differ by more than 1 year.
4-rounded to 1 decimal place to prevent small number counts in 2 variables missing columns.
5-Percent of 10,226 and does not add up to 100%, 5 most common combinations only. All other combinations of disagreement account for 2.75% of total.

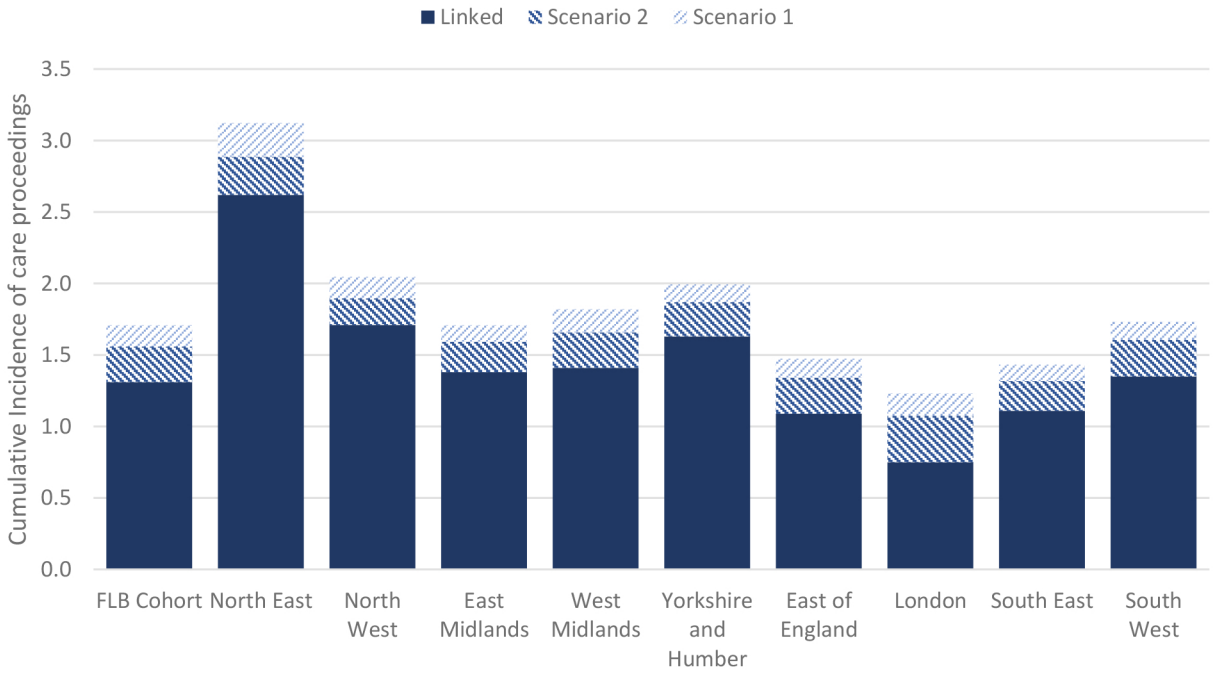
## 6. Sensitivity analysis of cumulative incidence of care proceedings at 10 years after a first live birth

Underestimation of cumulative incidence due to missed matches was explored using the records of mothers who did not link to HES.

Among unlinked mothers recorded only in Cafcass, year of FLB was estimated using maternal age at oldest child and first case start (where available). We excluded mothers with an estimated FLB before 2007, but included where could not be calculated (i.e. no maternal date of birth available). Under scenario 1, the least conservative estimate, all remaining mothers were added to the crude numerator of mothers who experienced CP within 10 years of a FLB. Under scenario 2, only mothers with a child less than 2 years at first case were added to the numerator, as we assumed that births close in time to care proceedings were more likely to have been in England. The percent increase in crude incidence from scenario 1 and 2 was applied to Kaplan-Meier estimates of cumulative incidence. These scenarios assume there are no duplicate mothers in the unlinked pool, that all children associated with the case were delivered within the NHS in England, all mothers without a date of birth were 15-50 years old at case start and that unlinked mothers would contribute similar follow-up time to linked women.

Scenarios 1 and 2 increased crude incidence of care proceedings to 1.4% (46,903/3,572,737) and 1.3% (42,803/3,572,737) respectively, representing an increase of 30.4% and 19.0%. Impact by region varied (Supplementary Figure 8), with the percent increase being greatest in London and the East of England (64.1% and 35.2% increase under scenario 1 and 43.4% and 23.1% under scenario 2). When the percent increases were applied to Kaplan-Meier cumulative incidence estimates for the FLB Cohort, incidence increased to 1.7% under scenario 1 and 1.6% under scenario 2.

Supplementary Figure 8: Cumulative incidence of care proceedings within 10 years of a first live birth as estimated from 1) mothers involved in care proceedings who linked to the first live birth cohort via Kaplan Meier analysis, 2) Kaplan Meier estimated cumulative incidence adjusted by percent increase in crude incidence via sensitivity scenario 1, and, 3) Kaplan Meier estimated cumulative incidence adjusted by percent increase in crude incidence via sensitivity scenario 2




Mothers included in Scenario 2 also contribute to mothers included in Scenario 1.