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. 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. 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. 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. 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. 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 1st April 2007 to 31st 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 1st April 1997 and 31st 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) [**?**], 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 1/5th 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+)

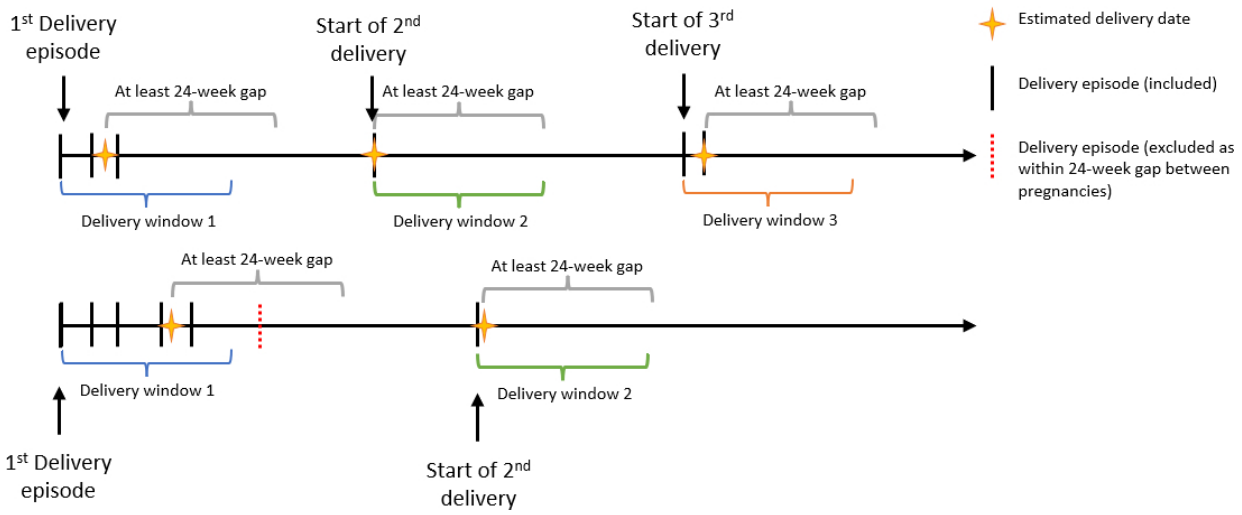
Once a date of delivery was estimated for delivery 1, the first episode associated with a second delivery could be identified, where available, for each woman. The subsequent delivery had to be at least 24 weeks after the first delivery date, and if episodes occurred before this time they were excluded.

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.

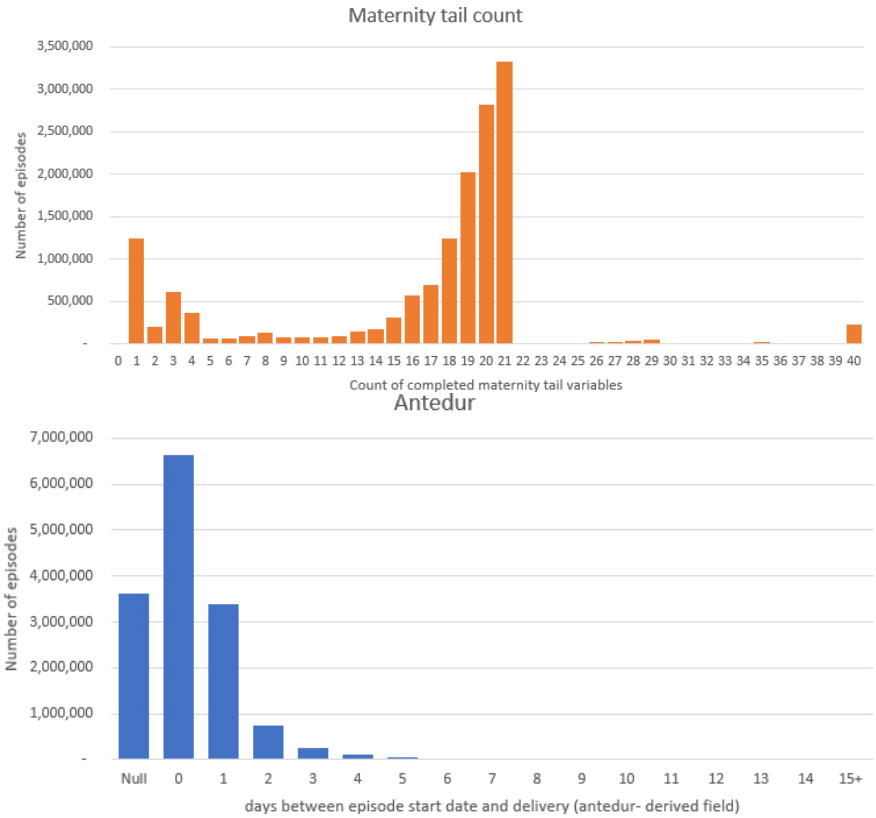
Supplementary Table 5: Criteria for identifying delivery episodes within Hospital Episode Statistics Admitted Patient Care data

Variable	**Variable name in HES APC**	**Criteria**	
Proceedural codes (OPCS-4.10)	*optern_1->optern_**	R14 – surgical induction of labour R15 – other induction of labour R17 – Elective caesarean delivery R18 – Other caesarean delivery R19 – breech extraction delivery R20 – other breech delivery R21 – forceps cephalic delivery R22 – vacuum delivery R23 – cephalic vaginal delivery with abnormal presentations of head at delivery without instruments R24 – normal delivery R25 – other methods of delivery R27 – other operations to facilitate delivery	OR
ICD- 10 diagnostic code	d*iag_1->diag_**	Z37 – outcome of delivery Z38 – liveborn infants according to place of delivery	OR
Delivery method	*delmeth_1*	0 = Spontaneous vertex (normal vaginal delivery, occipitoanterior) 1 = Spontaneous other cephalic (cephalic vaginal delivery with abnormal presentation of head at delivery, without instruments, with or without manipulation) 2 = Low forceps, not breech, including forceps delivery not otherwise specified (forceps, low application, without manipulation) 3 = Other forceps, not breech, including high forceps and mid forceps (forceps with manipulation) 4 = Ventouse, vacuum extraction 5 = Breech, including newlinetial breech extraction (spontaneous delivery assisted or unspecified) 6 = Breech 7 = Elective caesarean section 8 = Emergency caesarean section 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' N* = typically privately run providers 89997 = UK provider where no organisation code has been issued 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	

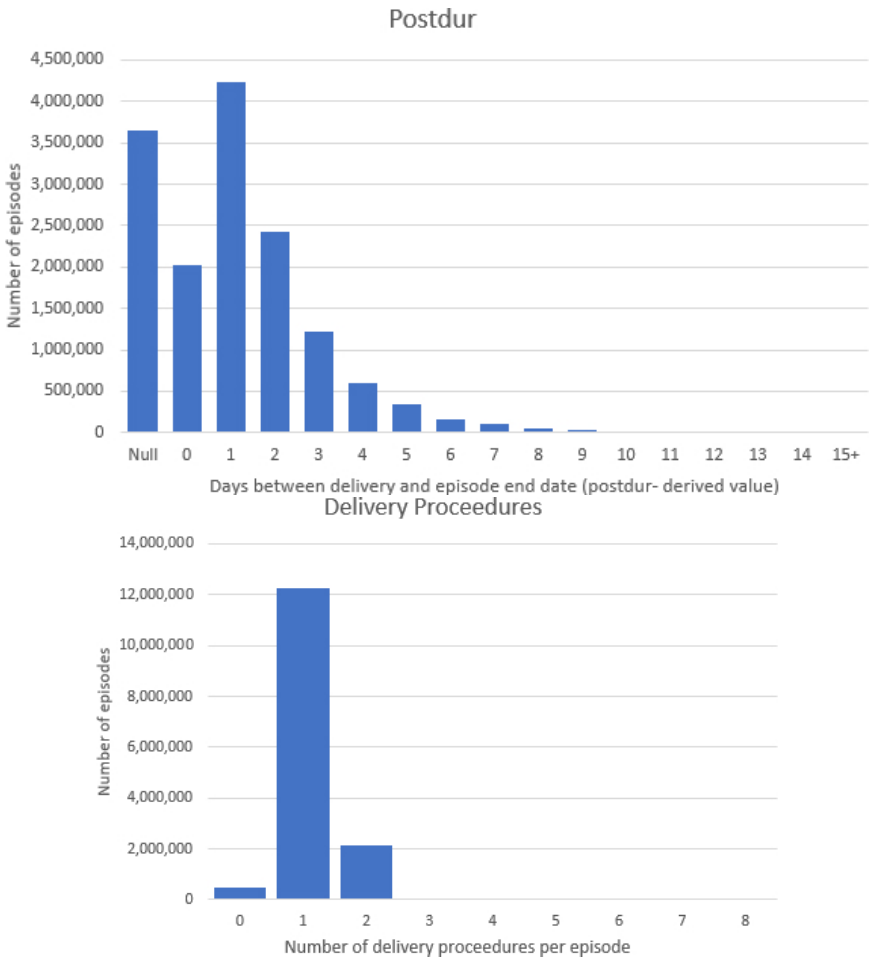
Supplementary Figure 1: An example timeline of how delivery episodes and estimated delivery date may look over a women's delivery episode history



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	≥ 3
Diag_*	Z37.0=single live birth Z37.1 = single still birth Z38.0=singleton, born in hospital Z38.1=singleton, born outside hospital Z38.2=singleton, not specified	Z37.2 = twins, both liveborn Z37.3 = twins, one liveborn one stillborn Z37.4 = twins, both stillborn Z38.3=twin, born in hospital Z38.4=twin, born outside hospital Z38.5=twin, unspecified	Z37.5 = other multiple births, all liveborn Z37.6 = other multiple births, some liveborn Z37.7 = other multiple births, all stillborn Z38.6=other multiple, born in hospital Z38.7=other multiple, born outside hospital 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) 3 = still birth (intrapartum) 4 = still birth (indeterminate)
diag_01–20	Z37.0 = single live birth Z37.2 = twins, both liveborn Z37.5 = other multiple births, all liveborn	Z37.3 = twins, one liveborn one stillborn Z37.6 = other multiple births, some liveborn	Z37.1 = single still birth Z37.4 = twins, both stillborn Z37.7 = other multiple births, all stillborn 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

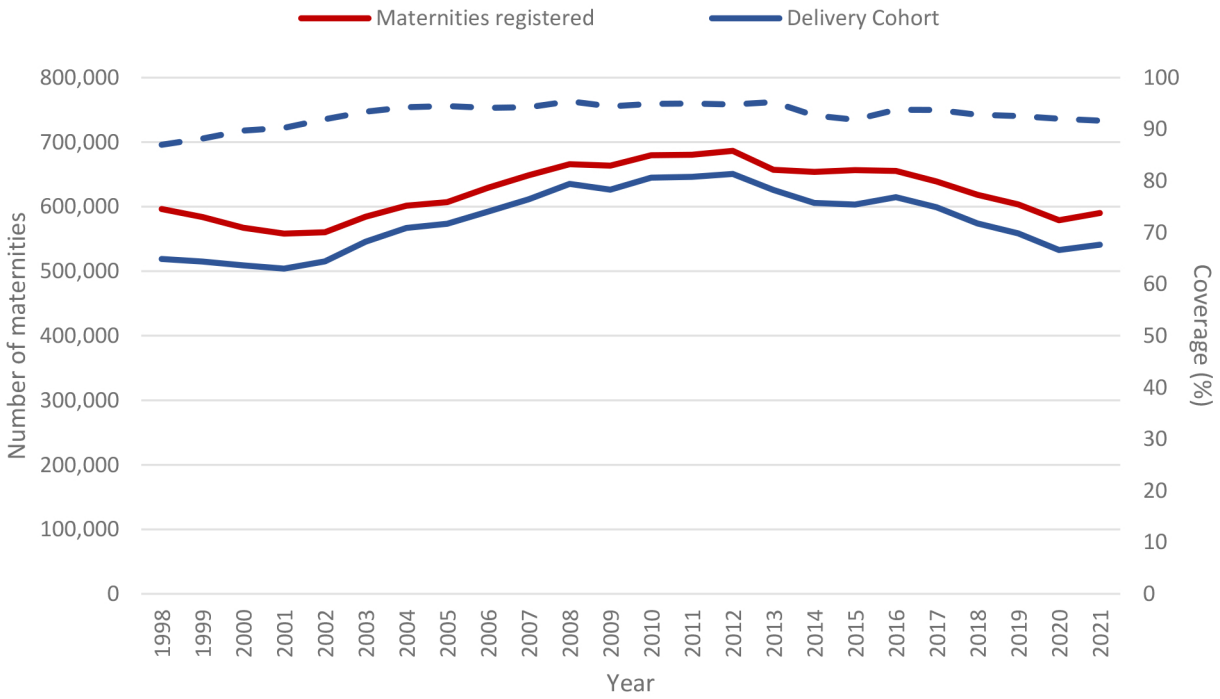
Women were included in the first live birth cohort if their first delivery record in the delivery cohort met the following criteria:

- 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

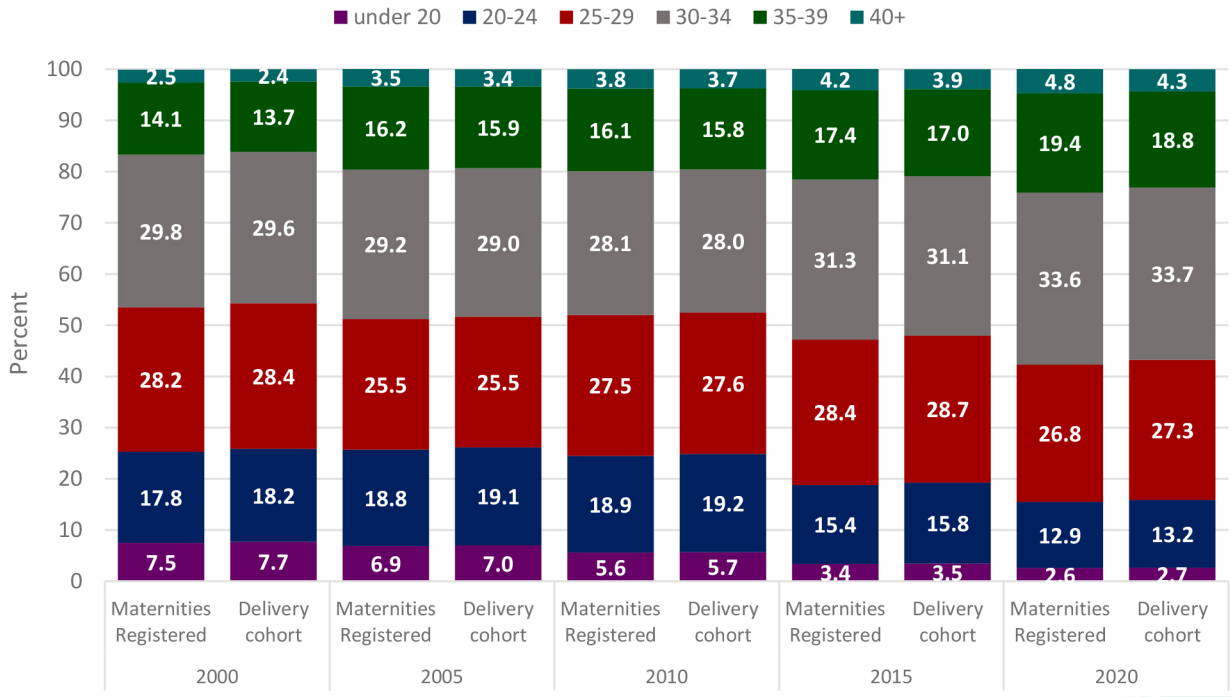
Demographic and delivery outcomes of mothers within the cohort were compared to civil registration of birth data for England, availabile from the Office for National Statistics (ONS), to determine whether the cohort was representative of all births in England (Supplementary Figures 3–7). To estimate 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)



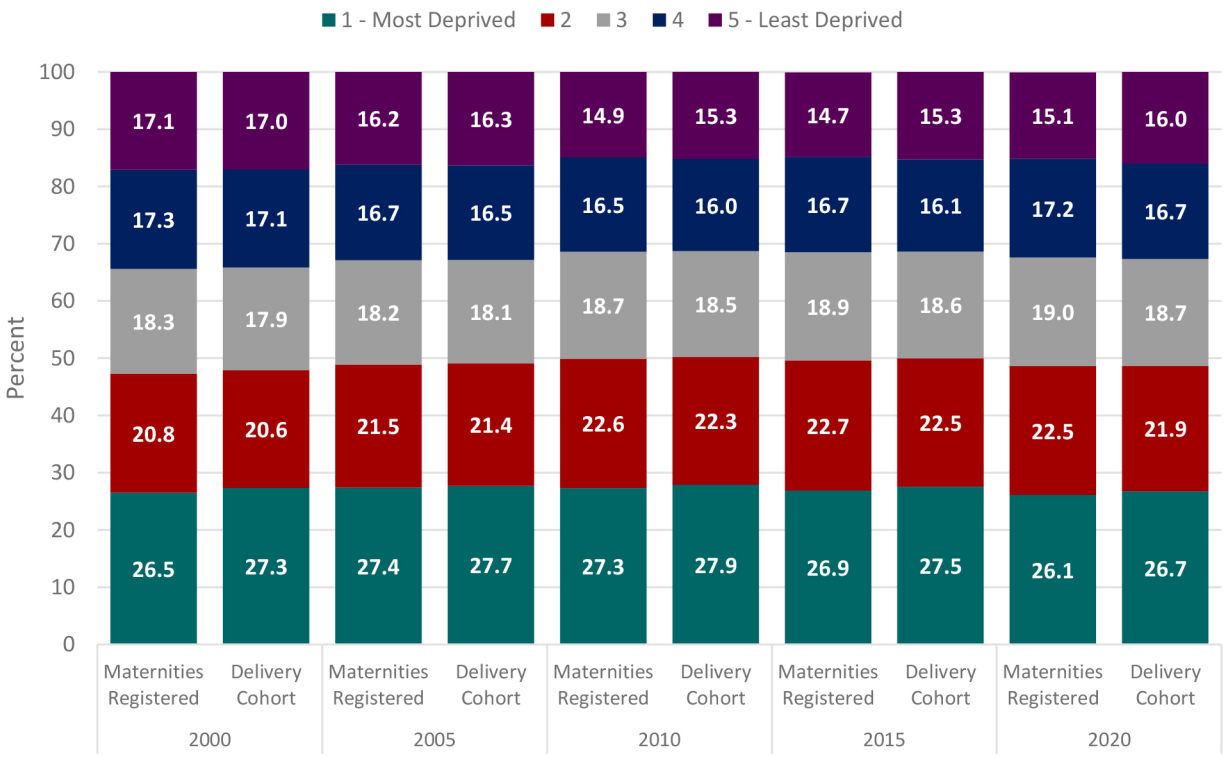
*delivery of live or still-born babies. If multiples were delivered only counted once.

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



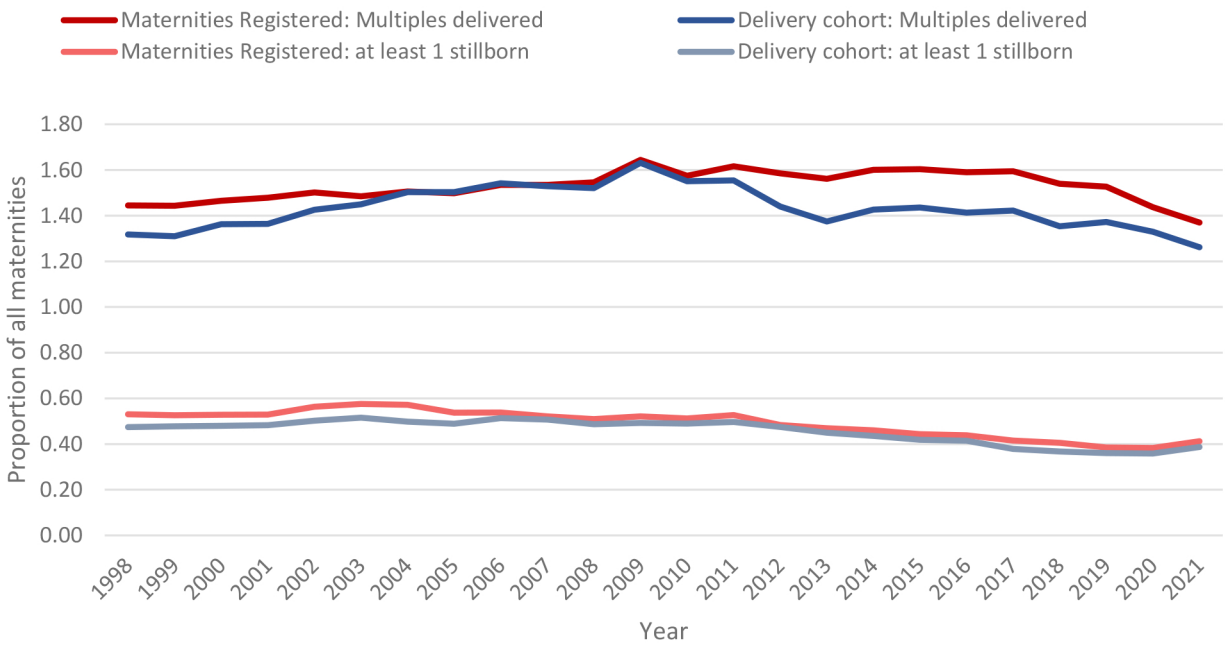
*delivery of live or still-born babies. If multiples were delivered only counted once.

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



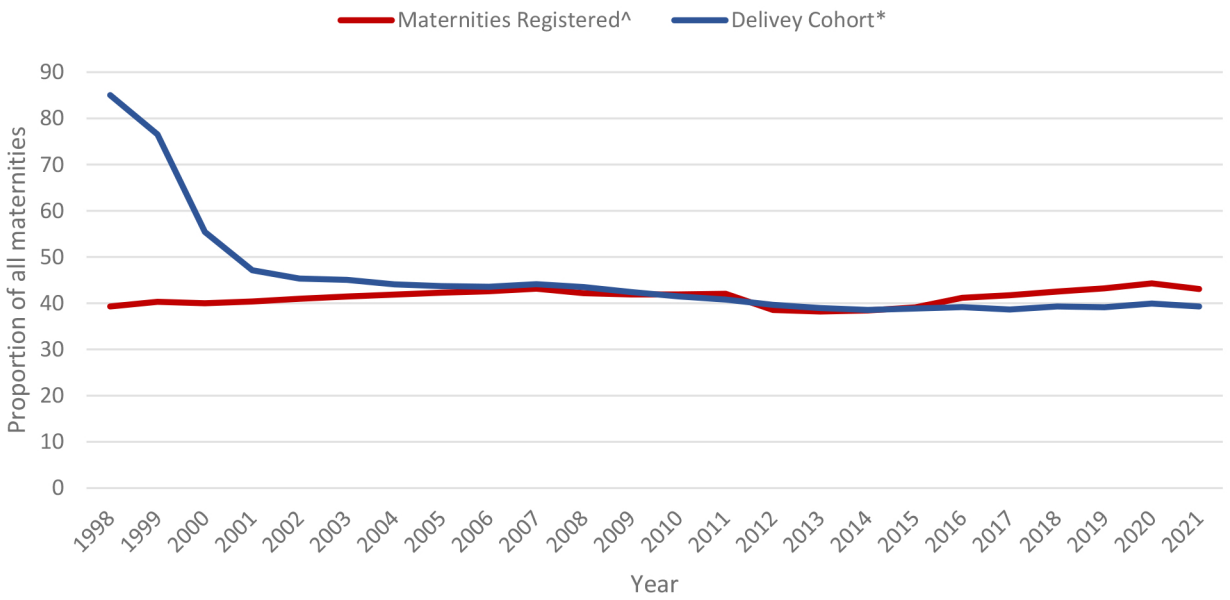
*delivery of live or still-born babies. If multiples were delivered only counted once.

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.

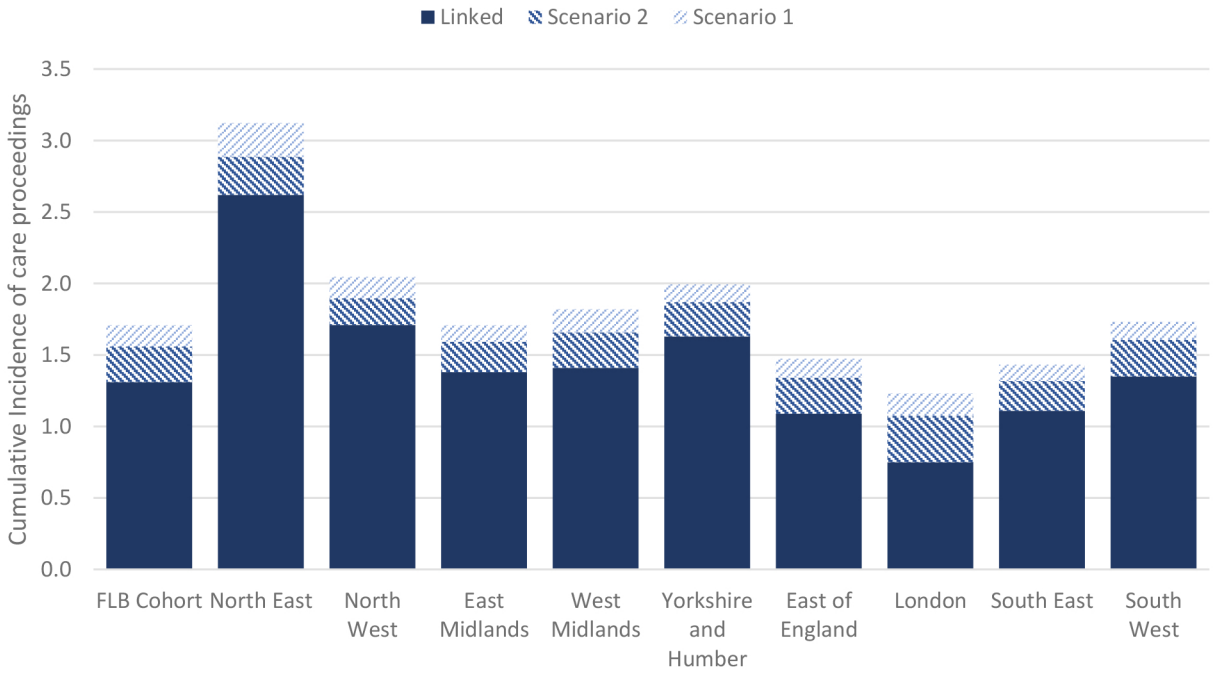
6. Sensitivity analysis of cumulative incidence of care proceedings at 10 years after a first live birth

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.

Temporary page!

LaTeX was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because LaTeX now knows how many pages to expect for this document.