

## Identifying the parental figures of children who enter care: The pros and cons of different data linkage methods using Welsh datasets

 Nell Warner<sup>1\*</sup>, Helen Hodges<sup>1\*\*</sup>, Jonathan Scourfield<sup>1</sup>, and Rebecca Cannings-John<sup>2</sup>

### Submission History

Submitted:	27/10/2025
Accepted:	19/01/2026
Published:	18/03/2026

<sup>1</sup> CASCADE, Cardiff University, Maindy Road, Cathays, Cardiff CF24 4HQ

<sup>2</sup> Centre for Trials Research, Cardiff University Health Park Campus, Neuadd Meirionnydd, Heath Park Way, Cardiff CF14 4YS

\*\*This author passed away in January 2026, prior to the publication of this manuscript

### Abstract

#### Introduction

Research on the parental figures of children who receive social care services is important, but there are challenges identifying parental figures, particularly fathers, from administrative data in some countries.

#### Objective

This paper compares methods for linking to parental figures using different datasets.

#### Methods

Children who entered out-of-home care in Wales between 2011 and 2019 were identified from the Looked After Children dataset for Wales (LACW). Those with Anonymous Linking Fields (ALFs) ( $n=10,679$ ) were linked to different datasets to identify parental figures. The Children and Family Court Advisory and Support Service (Cafcass Cymru) dataset was used to identify parental figures involved in public and private law cases. The Welsh Demographic Service Dataset (WDS) was used to identify parental figures resident with children before care entry and the National Community Child Health Database (NCCHD) was used to identify biological mothers. The 2011 and 2021 Censuses were used to identify if children were resident with parental figures on Census dates. Descriptive statistics were used to compare linkage rates to female and male parental figures across different datasets, the consistency of linkage and potential biases.

#### Results

NCCHD identified female parental figures for 86.9% of the children with an ALF. WDS identified female parental figures for 71.6% and male parental figures for 42.5%, while Cafcass Cymru data identified female parental figures for 53.9% and male parental figures for 44.0%. Response rates to the 2011 and 2021 Censuses were lower than for the general population, limiting their use.

#### Conclusion

NCCHD is useful for identifying biological mothers. WDS and Cafcass Cymru can both identify limited numbers of male parental figures, but the choice of linkage method would depend on the research questions to be answered.

#### Keywords

parents; mothers; fathers; child looked after

\*Corresponding Author:

Email Address: [warnerah@cardiff.ac.uk](mailto:warnerah@cardiff.ac.uk) (Nell Warner)



## Introduction

Understanding the population of children receiving social care services is important, especially because of their high levels of vulnerability and poor outcomes in adulthood [1, 2]. It is a larger population than many people are aware of, with the cumulative incidence by age 18 in England being 25.3% for child in need status and 6.9% for a child protection plan [3]. Children in state care, described as 'looked after' in UK law, are a group for whom local authorities have corporate parenting responsibility. The cumulative proportion for this group by age 18 in England has been identified to be 3.3% [4].

Understanding the circumstances of the parents of these children is key to understanding need and planning preventative services. In doing this, evidence about all parents and parental figures is important. Evidence about father figures is equally as important as that about mothers, given fathers' impact on children's development, potentially either positive or negative [5, 6]. However, much of the wider children's social care evidence base is dominated by information about mothers, with relatively little attention to father figures [7], partly because fathers are not very often engaged by services - a problem first noted well over 30 years ago [8], but recent evidence shows little sign of change [9].

While the need to include father figures in research is clear, when doing quantitative research, the availability of data about father figures makes this very challenging. Often, data are focused more on mothers. For example, UK datasets derived from survey data can fail to identify the differences between birth fathers and other forms of father such as stepfathers and foster fathers and often lack information about non-resident fathers [10].

The use of administrative data in children's social care research has been increasing in recent years [11], and it provides many advantages. These include its coverage of whole populations and scope for longitudinal analysis [12] and large sample sizes enabling subgroup analysis [13]. Research also benefits from its capacity to link together data from diverse sources [14]. This capacity has already been used to carry out studies linking information about children and parental figures, so that studies looking at the impacts of parental figures on children can be carried out [15–17].

There are however challenges for studies that link children and parental figures. Where birth data is available then this can be used to identify biological mothers but for father figures and non-biological mothers it is more challenging. A scoping review of international administrative data research identifying fathers in different ways was published in 2022 [18]. Most of the studies examined were carried out in Scandinavia, where personal identification numbers held on national population registers enable linkage between children and fathers to be carried out. In other countries where this facility is not available, alternative methods have been used. In Australia and Canada, paternal information was available on information relating either to health care numbers or birth registrations, while studies in both Taiwan and Spain had used information from health claims data to identify fathers.

In both the UK and the United States, methods using households or families based on address information have been used. This is done by providing different addresses with a code and then linking individuals with the same address code

together (see for example Rogers et al.'s research in Wales [19]). When no other sources of identifying parental figures are available, then the use of address information can provide details of the adults who children live with and who are playing a part in their lives. This has been used in studies to link household members and identify the effects mental health and substance misuse in the adults in households have on child outcomes including hospital admissions [20] and care entry [17]. This is an important strategy for identifying the people likely to have an impact on children. However, there are limitations, not least that we do not necessarily know how the individuals in the household are related to the child, nor do we have information about non-resident parental figures.

Information about non-resident parental figures is important for research relating to children receiving social care services, as these children are disproportionately likely to live in single-adult households [17]. The role that non-biological fathers play is also an important area for research, with examples over several decades of where practice has focused on mothers, but children have been seriously harmed or killed by father figures, including those without legal parental responsibility [9, 21].

When using linked administrative data, different datasets have the capacity to identify different types of parental figures. However, since these datasets have come from a range of sources the ability to link between them varies. This coupled with missing data about the parent-child relationship can produce bias. Depending on research questions, this could happen in a variety of ways, for example if data availability had changed over time, or varied according to the age of the child. Bias can also occur with father figures if data is only present for those in certain situations, such as resident fathers, those who have been involved in court proceedings or involved in services in some other way. In a study linking cohort study survey data to administrative data [22], it was found that the identification of fathers was affected by selection bias. There is currently a lack of published evidence comparing the relative benefits of the different linkage methods to identify parental figures, and the bias that might be introduced by each method.

This paper provides an assessment of the quality of the linkage of children who enter care in Wales to their parental figures using a variety of different UK administrative datasets. It relates to children who entered care between 1/4/11 and 31/3/19. This time frame was selected as it was carried out as part of a wider study on children returning home from care. More specifically it answers the following questions:

- What proportion of children who enter care can be linked to parental figures using different methods?
- Which types of parental figures can be identified using the different methods?
- Is there any consistency in the parental figures identified across datasets using the different methods?
- Is there any evidence of bias using different methods in relation to changes over time or according to the age of the child?

Note on gender: Throughout this paper we refer to the parental figures as "female parental figures" and "male parental figures." This is based on the "gender" variables provided in

the datasets we explored. These variables are referred to in the datasets as “gender” variables but rely on the way that individual’s gender has been recorded in the different dataset. Since we are not aware whether those recording the gender have consulted with individuals about how they identify they may therefore sometimes report on an individual’s sex rather than on the gender that those individuals identify as.

## Method

This was a data linkage study utilising administrative datasets relating to Wales, UK. These were accessed through the Secure Anonymised Information Linkage (SAIL) Databank, at Swansea University, which is an accredited trusted research environment [22]. The SAIL Databank contains a range of anonymised person-level datasets from different sources, including social care, health, education and justice.

### Identifying children in out-of-home care

Data regarding children who entered care and were placed outside their home between 1/4/11 and 31/3/19 were identified from the Looked After Children Wales dataset (LACW). This is a national dataset covering all children who are in local authority care in Wales. It is an annual return from local authorities to the Welsh Government [13]. To identify the first entry of children into care, rather than repeat incidences of recurrent cases, data from 1/4/02 was used so that a lookback period [23] of nine years could be excluded. LACW Children only receiving short breaks (under Part 6 section 76 of the Social Services and Well-being (Wales) Act 2014) were excluded, as were those who were in state care but placed at home with their family throughout, as this was a study of children who had been in out-of-home care returning to their families. Unaccompanied asylum-seeking children were also excluded, as they would not have families of origin in Wales. Information about the child’s age and year of first entry to care was also derived from LACW. LACW was linked to additional datasets using an Anonymous Linking Field (ALF) [24], to identify parental figures. All cases with an ALF were used, including those identified through fuzzy matching (with a probability >0.5). Previous work using the LACW data highlighted a relatively high proportion of children in the LACW dataset for whom an ALF could not be found, particularly in pre-school children. Because of this a six-step CLA matching algorithm has been developed to enhance ALF identification, which was utilised for this study [25]. This algorithm includes steps which identify children based on week of birth and gender, which could result in incorrectly identifying twins. Twins were therefore excluded from the study.

### Datasets used to identify parental figures

Information about the datasets used to identify parental figures is available in Table 1. More information about all these datasets is available on the Health Innovation Gateway website (<https://healthdatagateway.org/en>)

For each dataset, the proportion of children linked was assessed to determine if it was viable to identify parental

figures. For the purposes of this paper, parental figures are considered to be any individual who has been described either as a parent, or a stepparent, parent’s partner or adoptive parent, or an individual, at least 15 years older than the child, who was resident with them before care entry. Parental figures were identified for each dataset as follows:

#### NCCHD

Children were matched to their biological mothers using the maternal ALF field. This is a method that has previously been used to identify the mothers of children involved in care proceedings [26].

#### Cafcass (Children and family court advisory and support service) Wales

Cafcass Cymru data has previously been used for research on the parents of children involved in care proceedings [27]. Two methods for identifying parental figures were used for this study:

1. Using the Relationship Table: The Relationship Table shows how two different individuals found in the data are related to each other. The study children were identified in this table, and those related to them identified. The following relationships were used: Parent, Stepparent, Adoptive Parent, Mother’s partner, Father’s partner. Each of these was combined with the gender variable to create mother and father figures. Stepparents, and mothers’ partner and father’s partner were combined, and together with the gender variable were used to create a stepmother figure if the person was female, and a stepfather figure if they were male.
2. Identifying adults involved in different types of application: The second method used types of legal orders made in family courts, specifically those for care orders and for types of private law orders. When applications for these are made then those involved have specific roles. A care order is a court order to ensure that the local authority shares parental responsibility for a child with their parents [28]. For all the cases in which the local authorities have applied to courts to get a care order then those children should be recorded in the Cafcass Cymru data as subjects of those care order applications. Those who have parental responsibility for them prior to the care order applications should also be recorded as respondents to the care order applications. Care order applications for children who linked to the Cafcass Cymru datasets were identified and the applicants for those care order applications. Potential parental figures were identified by looking at the respondents to care orders [29]. To ensure that we are not using stepparents or kinship carers who were not responsible for the children before they first went into care it was decided to use the information from the child’s first care order only. Contact, residence and child arrangement orders were also identified. These are private law applications associated with individual’s contact or residency with children. Those who were either applicants or respondents to these

Table 1: Datasets used to identify parental figures

Dataset	Description	Used to Identify
National Community Child Health Database (NCCHD) <a href="https://healthdatagateway.org/en/dataset/360">https://healthdatagateway.org/en/dataset/360</a>	Database about the child health system in Wales, including birth registrations	Biological mothers
Cafcass (Children and Family Court Advisory and Support Service) Cymru <a href="https://healthdatagateway.org/en/dataset/328">https://healthdatagateway.org/en/dataset/328</a>	Information from Cafcass Cymru about children and other individuals who have been involved in family court processes in cases where children have been appointed Children's Guardians or court advisers.	Parental figures via the relationship table, respondents to care order applications, and applicants and respondents to Contact, Residence and Child Arrangement Orders
Welsh Demographic Service Dataset (WSDS) <a href="https://healthdatagateway.org/en/dataset/359">https://healthdatagateway.org/en/dataset/359</a>	Register of all individuals registered with a Welsh General Practitioner. Includes an anonymised address field	Individuals at least 15 years older than children who they were residing with before care entry
Office for National Statistics (ONS) 2011 Census Wales <a href="https://healthdatagateway.org/en/dataset/335">https://healthdatagateway.org/en/dataset/335</a>	Data relating to Wales from the 2011 Census	Family members
ONS 2021 Census Wales <a href="https://healthdatagateway.org/en/dataset/361">https://healthdatagateway.org/en/dataset/361</a>	Data relating to Wales from the 2011 Census	Family members

cases could be classified as parental figures, but only in cases that occurred before the child entered care. For parental figures identified through applications, the gender variable was used to identify if these were male parental figures or female parental figures. All but five cases had data relating to the gender of the person responding to the care order applications. Gender information from NCCHD/WSDS was used when this was missing.

## WSDS

The Welsh Demographic Service Dataset (WSDS) uses GP registrations to give households a Residential Anonymous Linking Field (RALF). This enables individuals registered at the same household to be linked (Rogers et al, 2009). The households that children were living in one week before they entered care were identified. This date was chosen as it was close to care entry as possible while giving some leeway in case the care entry date was mis-recorded. Taking an earlier date would have excluded more babies who may not have been alive and registered with a GP for sufficiently long before care. Defining households via RALFs can produce some households that may relate to institutions or that do not have adults in. Households were excluded if they included:

- More than ten people in them on a given date
- Had multiple groups of only teenagers living in them and children entering care on more than three different dates
- Had no individuals at least 15 years older than the children

For this dataset, parental figures were defined as all individuals living in the households who were at least 15 years older than the child. This age was chosen as linkage to NCCHD suggested that motherhood was very rare before the age of

15. The gender variable was used to identify if they were male adults or female adults.

## 2011 and 2021 censuses

Information about parental figures was obtained from both Censuses by looking at whether the children indicated that they were living with parental figures at the time of the Census. Children were classified as living with a parental figure if they had indicated that their relationship with any of the other household members was as a son, a daughter or stepchild.

## Analysis

Descriptive statistics were used to describe the levels of missing data and parental figures identified using each method. Where over 50% of the children could have at least one parental relationship identified, further analysis was carried out to check if ALFs were available for the parental figures. This level was chosen to provide enough parental figures for analysis to be carried out with the sample using one method of identifying parents only. While there is no universal cut off for this, it was decided for the purposes of the wider study to use 50% as a cut off, so that time resources were not used pursuing a method that would result in only a small sample. Bivariate analysis was carried out to explore the relationship between the identification of parental figures and the child's age group and year of first care entry. Consistency was explored by identifying ALF matches across datasets, and descriptive statistics used to present this. Because some datasets, such as Cafcass Cymru and WSDS, could potentially identify more than one parental figure of a particular gender, children were coded as having, for example, a female parental figure that matched across Cafcass and WSDS if any of the ALFs for any of the female parental figures found in Cafcass Cymru matched to any of the female parental figures found in WSDS. Where genders

differed across the datasets, then the gender from NCCHD was used if present, if not then the gender provided by WDS was used. All analysis was carried out using STATA 19.0.

## Results

### Proportions of children with relationships to parental figures identified

There were 13,823 children who entered out-of-home care between 1/4/11 and 31/3/19, identified from LACW dataset. Of the 13,823 cases, 10,744 cases had an ALF. Of these 65 were duplicated ALFs. Most duplications were caused by children entering care in more than one local authority, and a small amount were caused by children re-entering in the same local authority and being given a new identifier. Where this happened the first entry was kept, leaving a sample of 10,679 unique children.

The proportions of children with ALF availability by demographic characteristics are presented in Supplementary Table 1. The match rate is higher in children aged 2 and over and stays high throughout compulsory school years before dropping for those aged 16 and 17. There is no clear pattern relating to year.

Table 2 shows the numbers of children that could be linked to each of the different datasets, and for whom parental figures could be identified.

The match rates to the Cafcass Cymru dataset need to be considered within the context of how many children might be expected to have been involved with the family courts. Our analysis of the non-linked data (Supplementary Table 1) shows that 7,203 of the children have a careOrder at some point. This would suggest that parental figures were identified for 87.4% of these, and parental figures with an ALF for 79.8%.

The match rates to both the 2011 and 2021 Censuses were disappointing. We knew that a high proportion of the children would not have been born by 2011, however 6,995 children had been born and only 4,089 (58.4%) of these could be matched to the 2011 Census. By March 2021 all the children in the sample were born, and about 60.6% could be matched to the Census. Parental figures could only be identified for 2,504 (23.4%). Because of these low proportions we did not assess the numbers of parental figures with an ALF for the two Censuses.

### Types of Relationships Identified

Table 3 shows the numbers and percentages of children with male and female parental figures identified through different datasets. With respect to the Cafcass Cymru data a breakdown is provided to show how many of the parental figures were identified through each of the different Cafcass Cymru linkage methods.

NCCHD was limited to only biological mothers, however, for both the Cafcass Cymru and WDS datasets we could also assess the relative benefits of identifying male as well as female parental figures. Identification of individuals through the Cafcass Cymru relationship table also enabled us to identify the types of relationships that individuals had. Data from the Cafcass Cymru relationship table suggested that for

most children for whom female or male parental figures were identified, then it was the mother (99.7%) or father (98.5%) that could be identified. Some children had additional parental figures such as both a mother and a stepmother. While these insights made the relationship table useful, there were many children within the Cafcass Cymru dataset who did not link to it, and we do not know why. Of the three Cafcass Cymru methods trialled, identifying individuals through care order applications enabled both more female and male parental figures to be identified.

### Consistency across datasets

Figure 1 and Supplementary Table 2 present the findings of the exploration of consistency across the datasets. The figures in bold on the Venn diagram show the children who have at least one parental figure identified in different combinations of datasets. 37.7% of the children have a female parental figure identified in all three datasets, and for 3.5% of them a female parental figure could not be identified in any dataset. Identification of male parental figures was much lower generally with 34.4% of the children not having a male parental figure identified in any dataset.

The italic figures given in brackets present numbers of children for whom there were parental figures whose ALFs matched across different combinations of datasets. It shows that 33.1% of the children had a consistent female parental figure, with the same ALF who appeared in all three datasets. For male parental figures there were 1,471 children for whom the ALF of a male parental figure identified through Cafcass Cymru matched with that of an individual found in WDS.

### Assessment of bias

To assess the levels of bias in the datasets caused by any relationship between missingness and either the age of the children or year of care entry, bivariate analysis was carried out (See Table 4). For NCCHD the percentage of children with a parental figure goes down as the age of the child increases. For both Cafcass Cymru and WDS a higher proportion of children in the 3 to 6 group have parental figures identified than in the 0 to 2 group. For Cafcass Cymru this decreases quite a lot with time, whereas in WDS, there is a slight decrease in the number of female figures in the later age groups but an increase in male parental figures.

## Discussion

In this study we matched children who entered care to five different datasets covering Wales UK to assess the capacity of those datasets to identify both male and female parental figures. Linkage to both the 2011 and 2021 Censuses was poor and we did not pursue further identification of parental ALFs. Parental figures could be identified through the NCCHD, Cafcass Cymru and WDS, however there were few cases where the same parental figures could be found in all three. The likelihood of linkage was related to the age of the child. Identification of parental figures in Wales remains difficult and subject to bias when compared to countries with other

Table 2: Numbers and Percentages of Children matched to parental figures through different datasets

	Matched to Dataset	Matched to at least one parental figure	Matched to at least one parental figure with an ALF
	n (%)	n (%)	n (%)
NCCHD*	10521 (98.5)	9976 (93.4)	9278 (86.9)
Cafcass - Relationship Table	6892 (64.5)	5294 (49.6)	4973 (46.6)
Cafcass Cymru – care order Applications	6892 (64.5)	6296 (59.0)	5747 (53.8)
Cafcass Cymru -Private law Applications	6892 (64.5)	763 (7.1)	648 (6.1)
Any Cafcass Cymru Method	6892 (64.5)	not assessed	6207 (58.1)
WSD 1 week before care	8621 (80.7)	8144 (76.3)	8144 (76.3)
2011 Census	4089 (38.3)	3760 (35.2)	not assessed
2021 Census	6474 (60.6)	2504 (23.4)	not assessed

\*NCCHD = National Community Child Health Database, WSD = Welsh Demographic Service Dataset.

Table 3: Numbers and percentages of children with male and female parental figures identified through different datasets

	NCCHD*	Cafcass Cymru Relationship Table**	Cafcass Cymru care order Applications	Cafcass Cymru Private Law Applications	Any Cafcass Cymru method	WSD – 1 week before case
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Overall match to dataset	10521 (98.5)	6892 (64.5)	6892 (64.5)	6892 (64.5)	6892 (64.5)	8621 (80.7)
At least one female parental figure	9278 (86.9)	4562 (42.7)	5341 (50)	590 (5.5)	5757 (53.9)	7645 (71.6)
At least one male parental figure	0 (0)	3489 (32.7)	4051 (37.9)	608 (5.7)	4697 (44.0)	4539 (42.5)
Both female and male parental figures	0 (0)	3078 (28.8)	3645 (34.1)	550 (5.2)	4247 (39.8)	4040 (37.8)
Only female parental figures	9278 (86.9)	1484 (13.9)	1696 (15.9)	40 (0.4)	1510 (14.1)	3605 (33.8)
Only male parental figures	0 (0)	411 (3.8)	406 (3.8)	58 (0.5)	450 (4.2)	499 (4.7)
At least one parental figure	9278 (86.9)	4973 (46.6)	5747 (53.8)	648 (6.1)	6207 (58.1)	8144 (76.3)
Matched to dataset but no parental figures	1243 (11.6)	1919 (18.0)	1145 (10.7)	6244 (58.5)	685 (6.4)	477 (4.5)
Not matched to dataset	158 (1.5)	3787 (35.5)	3787 (35.5)	3787 (35.5)	3787 (35.5)	2058 (19.3)

\*NCCHD = National Community Child Health Database, WSD = Welsh Demographic Service Dataset, \*\*From Cafcass Cymru relationship table: Female parental figures reported - Mother = 4547, Stepmother/father's partner = 17, Adoptive mother = 20. Male parental figures reported - Father = 3,435, Stepfather/mothers' partner = 61, Adoptive father = 23.

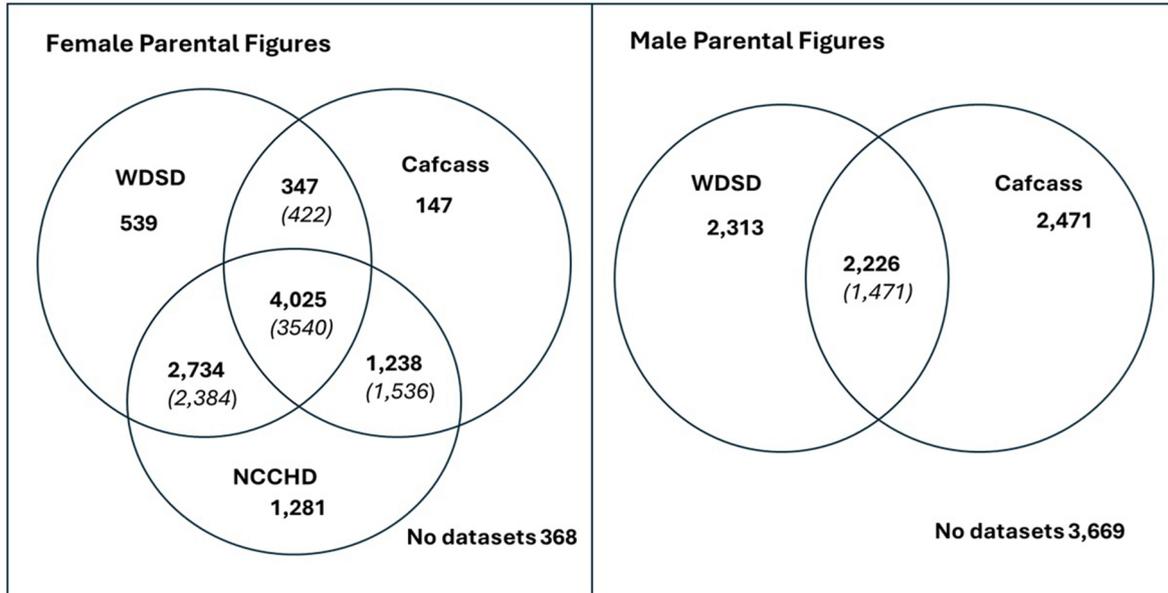
systems, such as the identification numbers held on national population registers found in Scandinavia.

The poor linkage rate to the two Censuses was disappointing. The 2011 Census data related to household composition on 27/3/11 and had an overall response rate of 94% [30]. Of the entire population in the 2011 Census, 91.4% had an ALF in SAIL. This means that around 85.9% of the resident population should be in the Census. Of the 6,995 children in our sample who were alive by then, only 4,089 (58.4%) of these matched to the 2011 Census. Some of those who entered care and who were alive on the Census date may not have been in Wales on that date. However, even bearing this in mind, it suggests that children who later entered care were considerably less likely to be living in households where the Census is completed than those who do not enter care. The ONS acknowledges that there are specific characteristics

associated with non-response of its Censuses which may result in bias [31] but there has not previously been any direct evidence about whether the households of children who enter care were less likely to complete the Census. However, there are characteristics associated with care entry that have also previously been associated with non-response bias, such as area level deprivation [32]. It may not be surprising if other family vulnerabilities related to the need to take children into care are also related to a likelihood of not completing the Census. It does mean that a reliance on information from the 2011 Census may be difficult for research designs in children's social care.

The 2021 Census had a response rate of 97%, however, ALFs were identified for a smaller proportion of the entire population (84.9%). We were able to find 6,474 (60.6%) of our children in the Census. Some of these would now be adults

Figure 1: Venn Diagram – Consistency of matching across datasets



Bold Figures indicate the number of children with parental figures identified in different datasets, italicised figures in brackets indicate parental figures whose ALFs match across different datasets.

Table 4: Numbers and percentages of children with parental figures identified by year of care entry and age group

	NCCHD	Cafcass Cymru		WDSO	
	Female parental figures n (%)	Female parental figures n (%)	Male parental figures n (%)	Female parental figures n (%)	Male parental figures n (%)
<b>Year of care entry</b>					
2011/2012	1085 (82.5)	614 (46.7)	477 (36.3)	945 (71.9)	583 (44.3)
2012/2013	1117 (84.7)	634 (48.1)	512 (38.8)	956 (72.5)	596 (45.2)
2013/2014	1085 (84.6)	625 (48.7)	510 (39.8)	908 (70.8)	532 (41.5)
2014/2015	1067 (85.8)	630 (50.6)	529 (42.5)	923 (74.2)	570 (45.8)
2015/2016	1146 (86.9)	737 (55.9)	592 (44.9)	937 (71.1)	566 (42.9)
2016/2017	1333 (89.3)	892 (59.7)	743 (49.8)	1052 (70.5)	602 (40.3)
2017/2018	1230 (90.4)	820 (60.3)	646 (47.5)	965 (71)	549 (40.4)
2018/2019	1215 (90.2)	805 (59.8)	688 (50.9)	959 (71.2)	541 (40.2)
<b>Age Group</b>					
0 to 2	2795 (92.1)	1805 (59.5)	1475 (48.6)	1307 (43.1)	741 (24.4)
3 to 6	2054 (91.2)	1560 (69.3)	1317 (58.5)	1891 (84)	1041 (46.2)
7 to 10	1567 (90.2)	1158 (66.6)	944 (54.3)	1502 (86.4)	858 (49.4)
11 to 14	1699 (81.3)	986 (47.2)	758 (36.3)	1731 (82.8)	1073 (51.3)
15plus	1163 (74.3)	248 (15.8)	203 (13)	1214 (77.6)	826 (52.8)
All	9278 (86.9)	5757 (53.9)	4697 (44)	7645 (71.6)	4539 (42.5)

NCCHD = National Community Child Health Database, WDSO = Welsh Demographic Service Dataset.

and would only appear if still living in Wales, so it is unclear how much of those missing might be due to this. The Census for England and Wales may therefore be a useful source of data for studies looking at outcomes for care experienced adults but, it was not a useful source for identifying parental figures of children who had previously entered care.

The other three datasets we explored were all better sources for identifying parental figures, but each had pros and

cons that might affect the types of study they are most suited to.

The dataset that was able to identify the highest proportion of parental figures was NCCHD. 86.9% of our sample could be linked to their biological mother with this dataset. There was a slight decrease in the likelihood of children being linked among older age groups. This could potentially be connected to the higher likelihood of these

children being born elsewhere and moving into Wales. NCCHD also has the advantage that we know how the parental figures are related to the children: they are the biological mothers, and this is clearly a key relationship in the child's life. In addition to the NCCHD, the SAIL databank also hosts the Maternity Indicators Dataset (MIDS) which also provides scope to link children to biological mothers. This was not used in this study since coverage of MIDS only starts in 2014 and many of the children in our sample would not have been born. However, in future this dataset maybe a useful tool to identify more biological mothers.

The obvious drawback with NCCHD is that fathers cannot be identified. Both Cafcass Cymru and WDSO data provide scope for identifying male parental figures, although both only identified male parental figures for less than 45% of the children in the sample. Whichever might be selected depends on the research questions being considered. The Cafcass Cymru relationship table provides data linkage to father figures for about 50% of the children who can be linked to the Cafcass Cymru dataset, while looking at respondents to care order applications enables father figures to be identified for 58.8%. These clearly show that even among those who have been involved with Cafcass Cymru there are still many children for whom father figures cannot be identified. Of the children who could be linked to Cafcass Cymru data, female parental figures were identified for 83.5%. This suggests that the poor linkage to father figures may reflect a lack of father figures, particularly those with parental responsibility, in the child's life. This ties in with analysis of Cafcass data from England which identified 20% of fathers as being missing from care proceedings [33]. The comparison of ALFs across households identified that of the children who matched to a female parental figure in Cafcass Cymru, 91.4% had identified their biological mother. It is also possible that some of the others were biological mothers but not identified in NCCHD. For 75.9% of the children with Cafcass Cymru female parental figures, that female parental figure could be identified in the WDSO dataset. Where this did not happen, it could be that the mother was not resident before care, but also because of data linkage problems. This contrasts with 31.3% of the Cafcass Cymru male parental figures who were found in the WDSO and found to be living with the children one week before care.

Cafcass Cymru, however could be a useful source of identifying parental figures for some studies, depending on the questions to be answered. Children appear in the Cafcass Cymru data if they have a care order application, so children who have been in care but only on a voluntary basis would be excluded from such studies, clearly creating a severe limitation of this method. However, arguably having a care order could be seen as a way of identifying a subgroup of children in care with a specific level of intervention need from the local authority, and for some studies it may be useful to consider only those with care order applications. It is the only method that we looked at that could identify non-resident fathers. We do not know how much of a role the non-resident parents may play in their child's life, and whether they have contact with them, so studies exploring impacts of this group will have to acknowledge this in their findings. The exploration of changes in the identification of parental figures via Cafcass Cymru data over time suggests this has been improving. If

this improvement continues, it may be a useful method for identifying more parental figures in the future. The likelihood of parental figures being identified did decrease according to the child's age. This may not be surprising given previous studies identifying that those who enter care on a voluntary basis as babies are much more likely than those who enter aged 10 plus to later receive a care order [34, 35]. This method is therefore less suitable for studies concerning the fathers of teenagers. Given all these differences and the intrinsic biases that they include, it will be important that any studies using this method define themselves as studies of children who at some point receive a care order, rather than children who are in care.

Overall, WDSO identifies at least one parental figure for a higher percentage of children than Cafcass Cymru, although the percentage of male parental figures identified by WDSO is quite similar overall. Even in the cases where male parental figures were identified across both datasets, only 66% matched to the same person. This is quite a plausible situation if, for example, a father with parental responsibility lives elsewhere, and a stepfather, or another adult man now lives with the child. Which is more useful to use in research studies will depend on the research questions being asked. There is a clear relationship between the identification of parental figures and the age group of the child, with parental figures being much less likely to be identified for the youngest age group. This is in part likely to be a consequence of children being taken into care at birth and therefore not being registered with a GP with their family before care entry.

This paper has clearly demonstrated that any of the methods used to identify parental figures will introduce bias into studies. This means that studies will need to be very clear to explore these issues. We particularly explored the age of the child and year of care entry, but it may also be that there are additional biases relating to additional issues such as the child's ethnicity, gender, level of deprivation, type of care experience and local authority which may affect outcomes and which we have not explored. We recommend that studies using linking to parental figures, carry out preliminary tests to explore for bias and incorporate methods to compensate for it, such as inverse probability weighting [36] if required.

## Limitations

This analysis was carried out using the LACW dataset which already suffers from having poor ALF match rate (See Table S1) which will create additional biases. To increase the sample, we used all children with an ALF, including those identified by fuzzy matching. This may have meant that some children could have been incorrectly identified. These will include biases created because of children being born outside Wales, and because of unstable NHS registrations. The study was based on children who entered out-of-home care and did not include children in care but placed with parents, nor children who were receiving support from local authorities but not in care, so we do not know how well our findings would extend to those populations. Our decision to identify children through the WDSO one week before care was taken to give a balance between getting a date that was as close as possible to care entry while still being a pre-care-entry house. However, it is

possible that some GP registrations may have been delayed or that individuals living with the children may not have been registered with the GP at that address. The decision to use individuals aged 15 years older to identify parental figures from WDS may have resulted in some individuals who were not truly parental figures having been categorised as parental figures. This is a limitation in terms of how parental figures are defined. Within the Cafcass data we excluded later care orders. While this was important to exclude individuals who only became involved in a child's life at a later time, it also inadvertently meant that it might have removed some genuine parental figures who might have been involved in the child's life earlier but did not appear in the administrative records until later. We did not pursue linkage through either Census, because of the poor linkage rates, however this may still provide some additional cases, and may work with other groups of children.

While this analysis has been able to highlight the relative benefits of identifying parental figures in different ways, its findings relate only to Wales, UK. Findings may also be of interest to those based in England, where some datasets are collected in a similar way, although in England there is currently no equivalent to the Welsh Demographic Service Data.

## Conclusion

This work has highlighted the relative value of using different data linkage methods to identify parental figures of children who enter care in Wales. The choice of method will depend on the research questions to be answered. Where only biological mothers are of interest then NCCHD can link to the greatest proportion of mothers. However, the importance of fathers for children's social care research cannot be understated. WDS and Cafcass Cymru provide the capacity to link to relatively similar numbers of father figures. WDS would be favoured for research where understanding the impact of resident father figures is important, whereas CAFCASS Cymru would be favoured if research can be limited to those who have at some time had a care order, and in which the influence of both resident and non-resident fathers is of interest. All the methods produce some bias in terms of the age group of children who are more or less likely to have parental figures identified and this will need to be considered in both research designs and the interpretation of findings.

## Acknowledgements

This paper is dedicated to the memory of our wonderful colleague and co-author Dr Helen Hodges who passed away in January 2026. Helen's unrivalled knowledge of administrative datasets was essential in planning the linkage used in this study and her reflections key to shaping the analysis. She will be sorely missed in the field of children's social care administrative data research.

We would also like to acknowledge the support of Health Care Research Wales, who funded the research project (SCF-22-02) and provide infrastructure funding for the CASCADE partnership.

## Ethics

This project was approved by the SAIL Databank Information Governance Review Panel.

## Data availability statement

The authors are unable to share the data that was used for this research. The data was accessed through the Secure Anonymised Information Linkage (SAIL) Databank, at Swansea University, which is an accredited trusted research environment. Use of the data was for this project only, and the authors do not have permission to share it.

## Declarations of conflicts of interest

None

## References

1. Batty GD, Kivimäki M, Frank P. State care in childhood and adult mortality: a systematic review and meta-analysis of prospective cohort studies. *Lancet Public Health*. 2022;7(6):e504–e514. [https://doi.org/10.1016/S2468-2667\(22\)00081-0](https://doi.org/10.1016/S2468-2667(22)00081-0)
2. Seker S, Boonmann C, Gerger H, Jäggi L, d'Huart D, Schmeck K, Schmid M. Mental disorders among adults formerly in out-of-home care: a systematic review and meta-analysis of longitudinal studies. *Eur Child Adolesc Psychiatry*. 2022;31(12):1963–1982. <https://doi.org/10.1007/s00787-021-01828-0>
3. Jay MA, Troncoso P, Bilson A, Thomson D, Dorsett R, Pearson R, De Stavola B, Gilbert R. Estimated cumulative incidence of intervention by children's social care services to age 18: a whole-of-England administrative data cohort study using the child in need census. *Int J Popul Data Sci*. 2025;10(1):2454. <https://doi.org/10.23889/ijpds.v10i1.2454>
4. McGrath-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;51:163–171. <https://doi.org/10.1016/j.chiabu.2015.10.020>
5. Sarkadi A, Kristiansson R, Oberklaid F, Bremberg S. Fathers' involvement and children's developmental outcomes: A systematic review of longitudinal studies. *Acta Paediatrica*. 2008;97(2):153–158. <https://doi.org/10.1111/j.1651-2227.2007.00572.x>
6. Scourfield J, Culpin I, Gunnell D, Dale C, Joinson C, Heron J, Collin SM. The association between characteristics of fathering in infancy and depressive symptoms in adolescence: a UK birth cohort study. *Child Abuse and Neglect*. 2016;58:119–128. <https://doi.org/10.1016/j.chiabu.2016.06.013>

7. Panter-Brick C, Burgess A, Eggerman M, McAllister F, Pruett K, Leckman JF. Practitioner review: Engaging fathers—recommendations for a game change in parenting interventions based on a systematic review of the global evidence. *J Child Psychol Psychiatry*. 2014;55(11):1187–1212. <https://doi.org/10.1111/jcpp.12280>
8. Parton C, Parton N. Women, the family and child protection. *Critical Social Policy*. 1988;8(24):38–49. <https://doi.org/10.1177/026101838800802403>
9. Davies J. *The Myth of Invisible Men. Safeguarding children under 1 from non-accidental injury caused by male carers*. London: Department for Education; 2021.
10. Goldman R, Burgess A. Where's the daddy? Fathers and father-figures in UK datasets. Contemporary Fathers in the UK series. Marlborough: Fatherhood Institute; 2017. Available from: <https://www.fatherhoodinstitute.org/contemporary-fathers-in-the-uk>.
11. Bailey G, Broadhurst K, Tranter K, Holmes L, Harron K, Hargreaves D, Woodman J, Griffiths LJ. *ADR England Community Catalyst: Children at Risk of Poor Outcomes, Full scoping review: Use of administrative data to understand children's involvement with children's statutory social care services*. London: Administrative Data Research UK; 2025.
12. Hurren E, Stewart A, Dennison S. New methods to address old challenges: The use of administrative data for longitudinal replication studies of child maltreatment. *Int J Environ Res Public Health*. 2017;14(9):1066. <https://doi.org/10.3390/ijerph14091066>
13. Allnatt G, Lee A, Scourfield J, Elliott M, Broadhurst K, Griffiths L. Data resource profile: children looked after administrative records in Wales. *Int J Popul Data Sci*. 2022;1(1). <https://doi.org/10.23889/ijpds.v7i1.1752>
14. Harron K, Dibben C, Boyd J, Hjern A, Azimae M, Barreto M L, Goldstein H. Challenges in administrative data linkage for research. *Big data & society*. 2017;4(2). <https://doi.org/10.1177/2053951717745678>
15. Johnson RD, Alrouh B, Broadhurst K, Ford D, John A, Jones K, Cusworth L, Akbari A, Smart J, Thompson S, Griffiths L. *Health vulnerabilities of parents in care proceedings in Wales*. London: Nuffield Family Justice Observatory; 2021.
16. Melis G, Bedston S, Akbari A, Bennett DL, Lee A, Lowthian E, Schlüter DK, Taylor-Robinson DC. Impact of Socioeconomic Conditions and Perinatal Factors on Risk of Becoming a Child Looked after: A Whole Population Cohort Study Using Routinely Collected Data in Wales. *Public Health*. 2023;224:215–223. <https://doi.org/10.1016/j.puhe.2023.09.001>
17. Warner N, Scourfield, J., Cannings-John, R., Rouquette, O.Y., Lee, A., Vaughan, R., Broadhurst, K. & John, A. Parental risk factors and children entering out-of-home care: The effects of cumulative risk and parent's sex. *Child Youth Serv Rev*. 2024;160:107548. <https://doi.org/10.1016/j.childyouth.2024.107548>
18. Lut I, Harron K, Hardelid P, O'Brien M, Woodman J. What about the dads? Linking fathers and children in administrative data: A systematic scoping review. *Big Data Soc*. 2022;9(1). <https://doi.org/10.1177/20539517211069299>
19. Rodgers SE, Lyons RA, Dsilva R, Jones KH, Brooks CJ, Ford DV, John G, Verplancke JP. Residential Anonymous Linking Fields (RALFs): a novel information infrastructure to study the interaction between the environment and individuals' health. *Journal of Public Health*. 2009;31(4):582–588. <https://doi.org/10.1093/pubmed/fdp041>
20. Paranjothy S, Evans A, Bandyopadhyay A, Fone D, Schofield B, John A, Bellis MA, Lyons RA, Farewell D, Long SJ. Risk of emergency hospital admission in children associated with mental disorders and alcohol misuse in the household: an electronic birth cohort study. *The Lancet Public Health*. 2018 Jun 1;3(6):e279-88. [https://doi.org/10.1016/S2468-2667\(18\)30069-0](https://doi.org/10.1016/S2468-2667(18)30069-0)
21. O'Hagan K. The Problem of Engaging Men in Child Protection work. *British Journal of Social Work*. 1997;27(1):25–42. <https://doi.org/10.1093/oxfordjournals.bjsw.a011194>
22. Jones KH, Ford DV, Thompson S, Lyons RA. A profile of the SAIL Databank on the UK secure research platform. *Int J Popul Data Sci*. 2019;4(2):1134. <https://doi.org/10.23889/ijpds.v4i2.1134>
23. Epping J, Geyer S, Tetzlaff J. The effects of different lookback periods on the sociodemographic structure of the study population and on the estimation of incidence rates: analyses with German claims data. *BMC Med Res Methodol*. 2020;20(1):229. <https://doi.org/10.1186/s12874-020-01108-6>
24. Ford DV, Jones KH, Verplancke JP, Lyons RA, John G, Brown G, Brooks CJ, Thompson S, Bodger O, Couch T, Leake K. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC health services research*. 2009 Sep 4;9(1):157. <https://doi.org/10.1186/1472-6963-9-157>
25. Bailey GA, Lee A, Ahmed S, Scanlon I, Cowley LE, Stuart A, Farr I, Brooks C, North L, Griffiths LJ. Improving opportunities for data linkage within children looked after administrative records in Wales. *International Journal of Population Data Science*. 2025 Feb 19;10(1):2383.
26. Farr I, Cowley L, Broadhurst K, Odd D, Jones C, Bailey G, Alrouh B, Abouelenin M, Cusworth L, Doebler S, Ford D, & Griffiths, L. Health service use of infants involved in family justice care and supervision proceedings in Wales: a data linkage study. *International Journal of Population Data Science*. 2024; 9(1):2362. <https://doi.org/10.23889/ijpds.v9i1.2362>

27. Bedston S, Philip G, Youansamouth L, Clifton J, Broadhurst K, Brandon M, Hu Y. Linked lives: Gender, family relations and recurrent care proceedings in England. *Child Youth Serv Rev.* 2019;105:104392. <https://doi.org/10.1016/j.childyouth.2019.104392>
28. Coram Child Law Advice. 2025. Care Orders, [Accessed Online 12/12/25: <https://childlawadvice.org.uk/information-pages/care-orders/>]
29. 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 Mar 26;5(1):1159. <https://doi.org/10.23889/ijpds.v5i1.1159>
30. Office for National Statistics. 2011 Census General Report for England and Wales. London: ONS; 2015.
31. ONS Characteristics of Census 2021 respondents by mode of completion, England and Wales. [Accessed online 12/12/25 at: [https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/articles/characteristicsofcensus2021respondentsbymodeofcompletionenglandandwales/2023-10-23?utm\\_source=chatgpt.com#main-points](https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/articles/characteristicsofcensus2021respondentsbymodeofcompletionenglandandwales/2023-10-23?utm_source=chatgpt.com#main-points)]
32. Goodman A and Gatward R. Who are we missing? Area deprivation and survey participation. *European journal of epidemiology.* 2008;23(6), pp.379–387.
33. Philip G, Bedston S, Youansamouth L, Clifton J, Broadhurst, K, Brandon, M. and Hu, Y, 'Up Against It' Understanding Fathers' Repeat Appearance in Local Authority Care Proceedings. 2021; Nuffield Foundation.
34. Cowley L, North L, Broadhurst K, Doebler S, Alrouh B, Cusworth L, Abouelenin M and Griffiths L. Born into care: Understanding care pathways and placement stability for infants in Wales. 2023. Swansea/Lancaster: Family Justice Data Partnership.
35. Edney C. What are the routes into care for young people in Wales? 2024;Swansea/Lancaster: Nuffield Family Justice Observatory.
36. Seaman, SR, and White, I. "Review of inverse probability weighting for dealing with missing data." *Statistical methods in medical research* 22.3 (2013): 278–295.



Table S1: ALF match rate by child characteristics

Year of First Care Entry	ALFs present		No appropriate ALF		All children n
	n	%	n	%	
2011_2012	1,315	77.3	386	22.7	1,701
2012_2013	1,319	75.4	430	24.6	1,749
2013_2014	1,283	75.7	412	24.3	1,695
2014_2015	1,244	73.4	450	26.6	1,694
2015_2016	1,318	78.0	372	22.0	1,690
2016_2017	1,493	82.5	317	17.5	1,810
2017_2018	1,360	78.3	377	21.7	1,737
2018_2019	1,347	77.1	400	22.9	1,747
<b>Age Group at Dirst Care Entry</b>					
Pre-birth to 2	3,034	56.0	2,385	44.0	5,419
3 to 6	2,251	91.9	198	8.1	2,449
7 to 10	1,738	94.5	102	5.5	1,840
11 to 15	2,829	92.3	236	7.7	3,065
16 or over	827	78.8	223	21.2	1,050
<b>Sex of the Child</b>					
Male	5,546	77.1	1,649	22.9	7,195
Female	5,133	77.4	1,495	22.6	6,628
<b>Disability</b>					
Disability indicated	724	91.5	67	8.5	791
Disability not indicated	9955	76.4	3077	23.6	13032
<b>Ethnicity</b>					
Asian	189	70.5	79	29.5	268
Black	142	64.0	80	36.0	222
Mixed	345	76.5	106	23.5	451
Other	109	68.6	50	31.4	159
White	9,621	78.0	2,706	22.0	12,327
Missing or inconsistent	273	68.9	123	31.1	396
<b>Care Order*</b>					
Care Order at some point	7,203	79.7	1,835	20.3	9,038
No careoOrder Recorded at any point	3,476	72.6	1,309	27.4	4,785
Total	10,679	77.3	3,144	22.7	13,823

\*Child had either a full care order or an interim care order at some point according to LACW data.



Table S2: Frequency of matches, and consistent matches across datasets

	<b>Total parental figures identified</b>	<b>Parental figures with matching ALFS</b>
	<b>n (%)</b>	<b>n (%)</b>
<b>Female Parental Figures</b>		
All three datasets	4025 (37.7)	3540 (33.1)
NCCHD & WSD	2734 (25.6)	2384 (22.3)
NCCHD & Cafcass Cymru	1238 (11.6)	1536 (14.4)
WSD & Cafcass Cymru	347 (3.3)	422 (4)
NCCHD only	1281 (12)	
WSD only	539 (5.1)	
Cafcass Cymru only	147 (1.4)	
no datasets	368 (3.5)	2797 (26.2)
<b>Male Parental Figures</b>		
WSD & Cafcass Cymru	2226 (20.8)	1471 (13.8)
WSD only	2313 (21.7)	
Cafcass Cymru only	2471 (23.1)	
None	3669 (34.4)	9208 (86.2)
Total	10679 (100)	10679 (100)

NCCHD = National Community Child Health Database, WSD = Welsh Demographic Service Dataset.

