Use of Linked Administrative Children's Social Care Data for Research: A Scoping Review of Existing UK Studies

G. Allnatt[1], M. Elliott[2], J. Scourfield[2], A. Lee[1] and L. J. Griffiths[1],*

[1]*Population Data Science, Swansea University Medical School, Swansea SA2 8PP, UK*
[2]*CASCADE, School of Social Sciences, Cardiff University, 1-3 Museum Place, Cardiff CF10 3BD, UK*

*Correspondence to L.J. Griffiths, Population Data Science, Swansea University Medical School, Swansea SA2 8PP, UK. E-mail: lucy.griffiths@Swansea.ac.uk

Abstract

Increasing access to children's social care data presents enormous potential for research and policy evaluation, with opportunities increased where data can be anonymously linked to other sources of information, such as health and education data. The purpose of this scoping review was to provide an overview of all UK data linkage studies that have used routinely collected individual-level children's social care administrative data. Six research databases were searched and twenty-five studies were identified as meeting the inclusion criteria, with the majority ($n = 18$) based on English data. Complexities and the time-consuming nature of these studies are highlighted, as are issues with missing data and inconsistencies in recording information across local authorities, impacting on the linkage process. Increased access to such data, and improvements to data capture, could improve the utility of these valuable administrative data assets in the social care sector.

Keywords: administrative data, Children's social care, data linkage, quantitative methods, review

BASW
The professional association for social work and social workers
www.basw.co.uk

Background

Information about children's social care services, including child welfare interventions is routinely collected by local authorities across the UK and submitted to Governments annually (Lee *et al.*, 2022; Emmott *et al.*, 2019; McGrath-Lone *et al.*, 2016). This administrative data provide information on children who receive any kind of support or intervention from the State and in some cases also those referred to children social care services where no further action was taken. Administrative data refer to information about persons or organisational activity collected routinely by Government other agencies for organisational purposes (Woollard, 2014). This micro or person-level data are very valuable for understanding the characteristics and outcomes of those receiving children's social care support.

It is increasingly recognised that the usage of administrative data for research is highly desirable where it can be achieved. Authors such as Brownell and Jutte (2013), Hurren *et al.* (2017) and Jay *et al.* (2017) make arguments in favour of using administrative data, that we summarise here. Sample sizes achievable are more representative and far larger than can be achieved with original data collection, facilitating population-level and sub-group analysis. Linking data to other relevant sources can significantly improve the reliability and accuracy of information. The data are not subject to reporting bias, such as stigma, that can apply to self-reporting certain kinds of contact with social services. Children's social care data can also provide information on interventions from welfare services. Data collection also tends to be longitudinal, allowing for change over time to be assessed. Finally, administrative data remove the burden on research participants to disclose sensitive issues.

These data-sets can be anonymised, reducing ethical difficulties with their re-use. There is also considerable potential for linkage to data from other sectors—for example linked education, health, family justice and social care data—which can establish certain outcomes (Johnson *et al.*, 2020; Cusworth *et al.*, 2021). This linkage can be done so that no individuals are identifiable by researchers using the linked data sets. For example, during the anonymisation process of data sources within the Secure Anonymised Information Linkage (SAIL) Databank, individuals are assigned an anonymised linking field (ALF) based on their NHS number, name, sex, date of birth and postcode. This process is carried out by a trusted third-party, meaning that data containing personally identifiable information are never held by the SAIL databank. ALFs can then be used to link person-level data-sets (Jones, 2019).

Data linkage is a growing field of enquiry in children's social care research and provides opportunities to link individual- and area-level administrative data and individual-level research data (e.g. surveys or trial

participation) across different sectors to answer complex research questions. Linking together information from large administrative or survey data-sets can provide researchers with opportunities to explore a range of issues and to address critical questions about the impact of child welfare interventions on individual and family lives over time (Edwards *et al.*, 2015). Similarly, McGrath-Lone *et al.* (2021) highlight that health, education and social care are strongly interconnected for children and young people, so record linkage offers a valuable opportunity to explore relationships between these interventions.

We set out to conduct a scoping review with two aims:

1. Provide an overview of all UK data linkage studies that have used routinely collected individual-level children's social care administrative data linked to other data sources. It focuses on the linkage of national, sub-national local and survey data.
2. Describe the benefits and limitations of using linked children's social care data and highlight gaps in the evidence base.

Although this is not a systematic review, we broadly followed the PRISMA (2020) checklist to guide our review and the reporting of results.

Governing legislation, data collection procedures and key terminology

Here, we introduce the legislative basis for children's social care in the four UK nations and the national administrative data sets in this field.

In England, the Children Act (1989) is the core governing legislation for children's social care services. The term children 'looked after' (CLA) refers to those children and young people who have been in the care of their local authority for more than 24 h. Children 'in need' refers to all those children who require support from children's services under section 17 of the Children Act 1989.

The SSDA903 statistical return, now known as the CLA Return, refers to the data collected and submitted to the Government annually by English local authorities. This contains information about CLA by local authorities and those who have recently left care. This data-set contains some demographic information such as date of birth, gender and ethnicity and extensive data on care histories. It does not contain the names of children. Instead, there is a unique child ID (local authority specific) assigned to the record and a pseudonymised unique pupil number (UPN) enabling linkage to the National Pupil Database (NPD) (McGrath-Lone *et al.*, 2016). The child ID can change following an adoption or a Special Guardianship Order. UPN is collected for linking to NPD specifically and UPN is collected for children of school-going age and some other specific groups—for example

children under school age if they have a statement of educational needs (SEN).

Similarly, the Children in Need (CIN) Census collects data on children and young people referred to local authority social care services because their health or development is at risk (Emmott *et al.*, 2019). This contrasts with Welsh data collection, where referrals into social care are not captured. The data contained within the CIN Census cover a wide range of CIN including those in care and those on child protection plans so there is an overlap with data held on the CLA Return.

In 2016, The Social Services and Well-being (Wales) Act 2014 replaced the Children Act (1989) as the core governing legislation for children's social care services in Wales. At the commencement of the Act, the term 'CIN' was replaced with the term 'Children Receiving Care and Support' (abbreviated as 'CRCS') to better describe the group of children and young people that receive help from social services. Terminology relating to CLA remained the same (Lee *et al.*, 2022).

Following the implementation of the new legislative framework in Wales, the data sources for capturing information on CRCS also changed. The Looked After Children Census replaced the existing SSDA903 data-set and the CRCS Census superseded the CIN Census.

In Scotland, the governing legislation for children's social care is the Children (Scotland) Act 1995. The Scottish Government collects data through the Children Looked After Statistics (CLAS data) and the Scottish Children's Reporter Administration (SCRA). The CLAS data-set holds information about the local authority, child details, episodes, placements and legal reasons for placements. The SCRA collects information about all children involved in the Children's Hearings System, including information about reasons for referral and child trajectory through the system.

The legislative framework for Northern Ireland's child protection system is set out in The Children (Northern Ireland) Order 1995. It sets out parental responsibilities, rights, duties and public authorities' powers to support children. Data about the support children receive from social care services are collected in the Child Order Statistical Returns data set annually. The information found in this data-set relates to CIN, children on the child protection register, looked after children and young people who have left care.

Methods

Search strategy

The following databases were searched: Medline, Web of Science, Applied Social Sciences Index Abstracts, Social Care Online, SCOPUS

and PSYCHINFO. Searches were conducted (by G.A.) on 16 April 2021. We used keywords that represented children's social care and data linkage, including: 'Children and social care data' OR 'children looked after' OR 'looked after children' OR 'children in need census' OR 'SSDA903' OR 'children receiving care and support' OR 'Scottish Exchange of data' OR 'Children's Social Care Statistics' AND (Link*). This was adapted for use with each database where necessary. We conducted additional searches in topic-specific websites of organisations involved with children social care data linkage research (The Rees Centre, What Works for Children's Social Care, Scottish Centre of Administrative Data, Administrative Data Research Wales and University College London (UCL)). In addition, we examined reference lists of relevant articles and reports.

Inclusion and exclusion criteria

Full eligibility criteria and rationale are given in Table 1. We included primary quantitative research, routinely collected individual-level administrative children's social care data, data linked to other data sources (administrative and research data), both national-, sub-national- and local-level studies and research published through peer-reviewed journals and research centre reports. We only included articles related to UK data sources on UK populations and published in the English language. No restrictions were set on year of publication.

Screening and data extraction

Two reviewers (G.A. and L.J.G.) screened all titles and abstracts of articles identified by the searches. Full-text articles were retrieved if potentially eligible. Characteristics of the included studies were extracted by one researcher (by G.A. then reviewed by L.J.G. and A.L.), including (1) citation; (2) sample size; (3) population; (4) social care data source; (5) linkage method; (6) other data sets used in conjunction with the children's social care data; and (7) summary of findings.

Risk of bias

Scoping reviews intend to map the concepts underpinning a research area and the primary sources and types of evidence available (Jolley, 2017) rather than assess the quality of individual studies. Study quality was therefore not appraised.

Table 1. Eligibility criteria and rationale

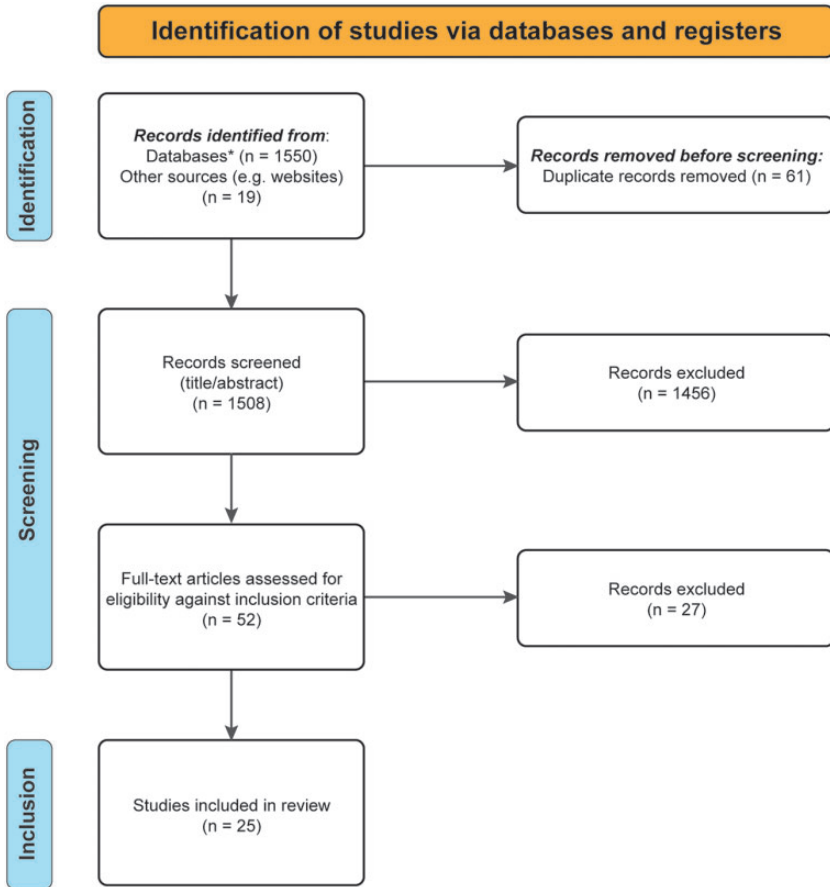| Criterion	Rationale
Primary quantitative research study	Primary research studies using individual-level administrative data published in peer-reviewed journals or identified in grey literature (i.e. via reports published on websites of selected research centres).
Included children's social care administrative data	Any studies that included national social care data sets, such as the CIN Census, CRCS Census, LAC Census or the same data fields sourced directly from local authorities; the following terms were used: 'CIN Census'; 'Children Receiving Care and Support'; 'Children Looked After'; 'SSDA903'; 'Scottish Exchange of data'; 'Children's Social Care Statistics'.
Included linkage to other administrative or research data.	Studies that linked children's social services data to other sector individual or area level administrative data, such as health and education, or research data (e.g. Avon Longitudinal Study of Parents and Children (ALSPAC)).
Use of national or local level social care data	Any use of a linked social care data-set as defined in inclusion criteria 2.
Conducted in any UK nation	UK studies only; use of data from all or any of the four nations.
Any study design	Any study design if it matches all the above inclusion criteria.
No date restrictions	Any published work if it matches all the above inclusion criteria.
Published in English	Only studies published in the English language.

Data synthesis and organisation

We summarise findings into the following broad themes:
- Geographical location of each study (England, Wales, Scotland and Northern Ireland) and whether the study was national or local.
- Type of data sources linked to children's social care data (health, education, deprivation measure and survey).
- Method of data linkage.
- Purpose of data linkage (results).

This review did not require ethical approval as it does not report on or involve any human subjects.

Results

We identified 1,569 articles in total—1550 from abstracts using electronic databases and 19 from other sources. After duplicates were removed, there were 1508 articles and reports, where the abstracts were reviewed for relevance. Fifty-two were identified as potentially eligible and full texts were retrieved. Of these, twenty-five met the inclusion criteria and

Figure 1: Flow diagram.

were included in the review (see Figure 1). The range of publication years for included studies is 2015–2021.

Table 2 summarises the details of the studies in this review.

The majority of studies ($n = 18$) were conducted in England (Sebba, 2015; Children's Commissioner, 2017, 2018, 2019 2020; Lugg-Widger, 2018; Baldwin, 2019, 2020; Sinclair, 2019, 2020; Teyhan, 2019; Berridge, 2020; Pearson, 2020; Webb *et al.*, 2020a,b; Zhang, 2020; Jay and Gilbert 2021). We found one study conducted in Wales (Elliott, 2020) and three in Scotland (Clark *et al.*, 2017; McMahon *et al.*, 2018; Hooper *et al.*, 2019). One study used a sample of local authorities in England and Wales (Dickens, 2019). Two studies involve a comparative analysis of the four UK nations (McCartan *et al.*, 2018; Bywaters *et al.*, 2020).

Table 2. Summary of studies (n = 25)

Study	Output type	Sample size and age	National or local level study	Country	Data-sets linked	Data-set type	Linkage process	Main results and conclusions
Studies covering whole UK								
Bywaters *et al.* (2020)	Published research article	N = 55 LAs; twenty-two from Wales, eighteen from England, eleven from N.I. and ten from Scotland. N = 12,412 on child protection plan or register (CP) N = 24,477 (CLA) (age 0–17)	Sub-national	UK	CLA; CP register and plans Deprivation indices; child population measures.	Individual-level social care data Area- and neigh-bourhood-level administrative data.	Small neighbourhood in which children lived, or from where they entered care, was used to link social care data to deprivation and population administrative data.	Structural inequalities in children's chances of being on a CP plan, register or being in out of home care, related to where the child's family live.
McCartan *et al.* (2018)	Published research article	N = 55 LAs; twenty-two from Wales, eighteen from England, eleven from N.I. and ten from Scotland. (age 0–17)	Sub-national	UK	CLA; CP plan Deprivation indices; child population measures.	Individual-level social care data Area- and neigh-bourhood-level administrative data.	Small neighbourhood in which children lived, or from where they entered care, used to link social care data to deprivation and population administrative data.	Extensive use of kinship care in the UK, particularly in areas of high deprivation.
Studies covering only England								
Jay and Gilbert (2021)	Published research article	N = 475,363 (age 5–16)	National	England	CIN; CLA NPD	Individual-level social care data. Individual-level school registration and attainment administrative data.	Pseudonymised UPN used to link social care data-sets to the NPD. Individual social care data-sets linked using anonymised child identifiers assigned by local authorities.	High proportion of CLA and CIN during school years in receipt of SEN provision. Overlap between the population served by social care and SEN provision in school.

(continued)

Table 2. (continued)

Study	Output type	Sample size and age	National or local level study	Country	Data-set information			Main results and conclusions
					Data-sets linked	Data-set type	Linkage process	
Baldwin *et al.* (2020)	Published research article	$N = 11,332$ (age 50–100 months)	Local	England	Children referred to child welfare services in a LA. Birth cohort study of expectant mothers of children born in a deprived multi-ethnic local authority.	Individual-level social care data. Individual-level questionnaire data.	A trusted third-party matched names, addresses and dates of birth to identify children who were in both data-sets.	Maternal smoking, mental ill health, lower maternal education level and other indicators of deprivation are salient risk factors for child maltreatment.
Berridge *et al.* (2020)	Report	$N = 471,688$	National	England	CIN; CLA NPD	Individual-level social care data. Individual-level school registration and attainment administrative data.	Pseudonymised UPN used to link social care data-sets to the NPD. Individual social care data-sets were linked using anonymised child identifiers assigned by local authorities.	Placement and educational instability linked with poorer attainment. Need for better visibility of the CIN group and the importance of early intervention.
Children's Commissioner for England (2017, 2018, 2019, 2020)	Report (annual series)	$N = 47,369$ (age 5–18)	National	England	CLA Social worker history School census via the NPD	Individual-level social care data. Individual child-level social worker history. Individual-level school registration and attainment administrative data.	Pseudonymised UPN used to link social care data-sets to the NPD. Individual social care data-sets were linked using anonymised child identifiers assigned by LA.	Approximately one in ten children in care experienced multiple placement moves; this number has remained unchanged since 2016. Seventeen per cent of CLA experienced school moves during the year (excluding moves from primary to secondary) and this has remained constant since 2016.

(continued)

Table 2. (continued)

Study	Output type	Sample size and age	National or local level study	Country	Data-set information			Main results and conclusions
					Data-sets linked	Data-set type	Linkage process	
Pearson *et al.* (2020)	Published research article	$N = 131$ Las	National	England	CLA HES APC	Individual-level social care data. Individual-level health data.	Aggregated values per local authority in both data-sets were linked according to the name of the LA.	Infant entry into care associated with increased prevalence of maternal history of adversity-related hospital admissions.
Sinclair *et al.* (2020)[a]	Published research article	$N = 642,805$ (age 0–16)	National	England	CLA NPD	Individual-level social care data. Individual-level school registration and attainment administrative data.	Pseudonymised UPD used to link the CLA records to the NPD.	CIN with earlier and persistent needs had lower educational attainment than other groups. Educational disadvantage of CLA mostly predates being in care.
Webb *et al.* (2020a)	Published research article	$N = 52,179$ (age 0–17)	National	England	CIN; CLA; CP Plan Index of multiple deprivation; estimates of child population and ethnic density; estimates of infant mortality; Job seeker's allowance claimant rate; educational level of population and income inequality before housing costs	Individual-level social care data. LSOA-level administrative data. LA-level aggregated administrative data.	LSOA code in which children lived, or from where they entered care, used to link social care data to LSOA-level administrative data. LA information within the social care data-sets used to link to LA-level administrative data.	Differences in ethnic group interventions depending on levels of deprivation. Socioeconomic inequalities in CP contingent on the ethnicity of the population.
Webb *et al.* (2020b)	Published research article	$N = 860,000$ (age 0–17)	National	England	CIN; CLA; CP plan Index of multiple deprivation; income inequality before housing costs	Individual-level social care data. MSOA-level administrative data.	MSOA code in which children lived, or from where they entered care, used to link social care data to MSOA-level administrative data.	Low overall deprivation and high-income inequality associated with greater inequalities in neighbourhood intervention rates.

(continued)

Table 2. (continued)

Study	Output type	Sample size and age	National or local level study	Country	Data-set information			Main results and conclusions
					Data-sets linked	Data-set type	Linkage process	
Zhang *et al.* (2020)	Published research article.	N = 14,062 (age 21–73 months)	Local	England	CP Register ALSPAC	Individual-level social care data. Individual-level birth cohort.	Children from the social care data-set matched to children in the birth cohort using personally identifiable information. Matched records then anonymised for analyses.	Risk of emotional and behavioural problems higher in children whose mothers have had contact with social workers.
Sinclair *et al.* (2019) [a]Secondary report from initial study by Sebba *et al.* (2015)	Published research article	N = 642,805 (main cohort) N = 6,236 (CLA) N = 20,383 (CIN)	National	England	CIN; CLA NPD	Individual-level social care data Individual-level school registration and attainment administrative data.	Pseudonymised UPN used to link the CLA and CIN records to the NPD.	Care can benefit the educational attainment of those who enter it before the age of 14 years. CIN who remain at home can fall behind their peers.
Baldwin *et al.* (2019)	Published research article	N = 363 (age 0–7)	Local	England	Child welfare data-set Survey of social workers; interviews with children's caregivers	Individual-level social care data Individual-level qualitative data	Exact matching process not detailed; likely that social care data-set was matched to interview and survey data using personally identifiable information.	Higher odds of a child in out of home care having a mental health problem and a reactive attachment disorder.
Teyhan *et al.* (2019)	Published research article	N = 12,868 (age 0–18)	Local	England	CIN; CLA NPD ALSPAC	Individual-level social care data Individual-level school registration and attainment administrative data. Individual-level birth cohort.	Previous work provided a link between the ALSPAC collection and NPD. NPD records then linked to a limited number of CIN and CLA records via a pseudonymised UPN.	CLA and CIN had worse educational outcomes relative to their peers.

(continued)

Table 2. (continued)

Study	Output type	Sample size and age	National or local level study	Country	Data-set information			Main results and conclusions
					Data-sets linked	Data-set type	Linkage process	
Lugg-Widger et al. (2018)	Published research article	N = 1,545 (aged under 6 years old)	Local	England	CIN; CLA	Individual-level social care data	Maternal self-report data linked to HES data by NHS number, date of birth, postcode and gender for both mothers and children. NHS number alone was used to match to mortality records. Further matching to the NPD used first name, surname, date of birth and postcode. CIN and CLA data linked to the NPD using UPN.	Study found support for the principle of trial data linkage in the context of child maltreatment. High levels of matching accuracy to health and NPD for a sample of mother–child dyads.
					NPD	Individual-level school registration and attainment administrative data.		
					Maternal outcomes; GP consultations; immunisations; abortions; inpatient; outpatient; A&E attendance (HES)	Individual-level event-based health record data.		
					Mortality records	Individual-level administrative data.		
					Maternal self-reporting	Individual-level pregnancy and childbirth trial participant data.		
Bywaters et al. (2018)	Published research article	N = 1,432,180 (age 0–17)	National	England	CIN; CLA	Individual-level social care data	Small neighbourhood in which children lived, or from where they entered care, used to link social care data to deprivation and population administrative data.	Relationship identified between deprivation and intervention rates and significant inequalities between ethnic categories.
					Child population estimates; index of multiple deprivation scores	LSOA- and local authority-level administrative data.		
Sebba et al. (2015)	Report	N = 622,970 (aged 15 years on 12 September 2012)	National	England	CLA	Individual-level social care data	Pseudonymised UPN used to link the CLA records to the NPD.	CIN did least well academically. SEN common amongst the CLA population.
					NPD	Individual-level school registration and attainment administrative data.		

(continued)

Table 2. (continued)

Study	Output type	Sample size and age	National or local level study	Country	Data-sets linked	Data-set type	Linkage process	Main results and conclusions
Studies covering England and Wales								
Dickens et al. (2019)	Published research article	N = 616 (age 0–17)	Local	England and Wales (sample from six LAs 5× England 1× Wales)	CIN; CLA (England and Wales) Social work case files[a] [a]For a subsample of children Cafcass database	Individual-level social care data Individual-level social care data Individual-level family justice system data	Exact matching process not detailed; likely that social work case files were matched to the Cafcass database by personally identifiable information. Unclear how the CIN and CLA data-sets were linked to other included data-sets.	Interactions identified between bringing care proceedings and demand for care placements. Decrease found in cases ending with adoption plans and an increase in cases ending with children living with relatives and children returning or remaining with parents.
Studies covering only Wales								
Elliott (2020)	Published research article	N = 8,853 (age 0–17)	National	Wales	CLA data Child population estimates; Welsh index of multiple deprivation (WIMD)	Individual-level social care data LSOA-level administrative data.	LSOA from where children entered care, used to link social care data to administrative data.	Children most disadvantaged neighbourhoods more likely to enter care than those living in the least deprived. Children entering care do so disproportionately from the poorest areas. Increase in children entering care since 2007.
Studies covering only Scotland								
Hooper et al. (2019)	Report	N = 1,834 (CLAS) N = 1,396 (SCRA) (age 0–5)	National	Scotland	CLA Statistics (Scotland) Scottish Children's Reporter Administration	Individual-level social care data Individual-level family justice system data	Data-sets linked using child date of birth, gender and LA at the time children became looked after.	Study highlights limitations of linkage process. However, match rate of records for 1,000 children from a sample of 2,230; match rate varied across local authorities.

(continued)

Table 2. (continued)

Study	Output type	Sample size and age	National or local level study	Country	Data-set information			Main results and conclusions
					Data-sets linked	Data-set type	Linkage process	
McMahon *et al.* (2018)	Published research article	$N = 10,924$ (age 5–11)	National	Scotland	CLA statistics (Scotland) Pupil census (Scotland)	Individual-level social care data Individual-level school registration and attainment administrative data.	Scottish candidate number (SCN) used to link children between social care and education records. Child date of birth, gender and home postcode from the Pupil Census used to link to all health data-sets.	School-age CLA have a history of greater dental health needs, higher hospital admissions levels for dental extractions and insufficient access to regular dental services.
					Scottish morbidity records (SMR01) hospital discharge; national dental inspection programme	Individual-level event-based health record data.		
Clark *et al.* (2017)	Published research article	$N = 637,815$ (age 4–19)	National	Scotland	CLA statistics (Scotland) Pupil census (Scotland)	Individual-level social care data Individual-level school registration and attainment administrative data.	SCN used to link children between social care and education records. Child date of birth, gender and home postcode from the pupil census were used to link to the community health index database.	Variation in completeness of individual records across local authorities which impacts on linkage. CLA have poorer health outcomes than their non-looked after counterparts.
					Community health index database	Individual-level event-based health record data.		

[a]Studies are grouped by the country that they use data from and then ordered by descending publication year within each group.

We were particularly interested in exploring the types of data sources used to link children's social care data, match rates and any associated challenges. We found that education data sources, namely the NPD, were linked to children's social care data in most of the studies ($n = 11$) (Sebba *et al.*, 2015; Children's Commisioner, 2017, 2018, 2019, 2020; Lugg-Widger *et al.*, 2018; Teyhan *et al.*, 2019; Sinclair *et al.*, 2019; Berridge *et al.*, 2020; Sinclair *et al.*, 2020; Jay and Gilbert, 2021). This linkage is relatively straightforward in England as the social care data-sets sit within NPD. These studies found that whilst children and young people in care did have low educational attainment, this could not be solely attributed to their involvement with the care system. Sinclair *et al.* (2020) found that lower than average educational attainment for CLA and CIN likely reflects early environment, special educational needs and poor relationships with secondary school.

Four of the studies linked routinely collected children's social care data to health data sources (Clark *et al.*, 2017; Lugg-Widger *et al.*, 2018; McMahon *et al.*, 2018; Pearson *et al.*, 2020). For example, Pearson *et al.* (2020) use the individual-level CLA return and the Hospital Episode Statistics Admitted Patient Care (HES APC) data set to calculate local authority-level rates of infant entry to care and indicators of maternal adversity (such as substance misuse, exposure to violence and mental health problems) respectively. Using these area-level aggregate data, the researchers established significant variation in infant entry into care depending on maternal adversity. However, because it was not possible to link hospital and CLA data at an individual-level, due to the lack of a common pseudonymised identifier in the data-sets, they could not examine whether children born to women with a history of maternal adversity-related hospital admissions were more likely to be placed into care during infancy. In another study, McMahon *et al.* (2018) linked dental records with children's social care data. They found that CLA are more likely to have dental treatment needs and less likely to access preventative dental services.

Five of the studies ($n = 5$) linked children's social care data to individual level research data which included questionnaire and trial participation data (Baldwin *et al.*, 2019; 2020; Lugg-Widger *et al.*, 2018; Teyhan *et al.*, 2019; Zhang *et al.*, 2020). Baldwin et al (2019) used administrative child welfare data linked to primary data from a survey of social workers and interviews with care givers. They found a higher rate of children in out of home care had a reactive attachment disorder compared with those who had never been in care. Two of the studies used the ALSPAC cohort study data-set, linked to routinely collected children's social care data (Teyhan *et al.*, 2019; Zhang *et al.*, 2020). Zhang *et al.* (2020) studied the scale of false reporting of social worker contact and found high incidences of false-negative reporting. They suggested several

reasons for this, including errors in the linkage process and the stigma attached to having a social worker, particularly for child maltreatment.

Five (*n* = 5) of the studies linked children's social care data to an area-level deprivation index (Bywaters *et al.*, 2018, 2020; McCartan *et al.*, 2018; Elliott, 2020; Webb *et al.*, 2020a).

Match rates were recorded in the majority of studies. For example, Lugg-Widger *et al.* (2018) matched 99.9 per cent of records from health, education and social care data-sets. However, others report the need for caution when reporting results from data linkage studies. Hooper *et al.* (2019) note that the process for data linkage in children's social care data was complex and time-consuming. They found that the records did not match as expected in just under a third of their sample, with many inaccuracies in the linking process. This had implications for the extent to which the data provided reliable and accurate information.

When considering the ethical implications of data linkage studies, it is important to note that teams using data linkage for research purposes do not have access to individual identifying information. They are typically provided with data-sets that are stripped of personal information replaced with pseudonymised identifiers that have been used to match records from across data sources. For example, McMahon *et al.* (2018) and Hooper *et al.* (2019) report use of pre-linked data-sets which were fully anonymised, linked by a trusted third party and analysed within a safe haven. Whilst not all studies reported the linkage source, we assume study teams used a trusted third party to undertake the data linkage as per standard procedure.

However, it is acknowledged that access to such personal information about individuals can raise questions about confidentiality and the potential to identify them. Enhanced governance will ensure that work in this field can continue without raising confidentiality concerns and protect the identity of all vulnerable children and families involved in children's social services.

Discussion

To our knowledge, this is the first review of UK studies linking children's social care administrative data with other data sources. This review identified twenty-five studies based on routinely collected social care administrative data linked to other administrative data such as health and education records, or to research data (survey, trial and cohort study). Whilst only a scoping review, we applied systematic review principles as far as possible and developed a thorough search strategy of peer-reviewed publications. We also carried out searches of reference lists and grey literature from social care research centres. Twenty-five studies are an encouraging number considering that linked children's social care

data for research purposes are a relatively new area of enquiry and considering the limited (although growing) access that researchers have to these data. This review provides evidence on the current standing of the data linkage process, highlighting the benefits and challenges of conducting data linkage research in children's social care.

Firstly, the review has shown that data linkage can be used for research to explore complex child welfare issues and contribute to the quantitative children's social care research portfolio. This is evidenced by the ($n = 25$) studies presented in Table 2. The studies show the scope of social care issues that can be explored using data linkage and highlight the benefit of interdisciplinary research. The linkage of population-level data-sets to other administrative data allowed researchers to build a partial longitudinal picture of the experiences of vulnerable children, facilitating a greater understanding of the long-term implications of adverse childhood experiences and social services contact for health and education outcomes (Lee *et al.*, 2022). Data linkage studies can also influence the direction of UK children's social care policy and practice (McGrath-Lone *et al.*, 2016).

We also sought to review how research teams in the included studies 'navigated' the matching process and whether the match rates presented were sufficient to allow for analysis and infer results. All studies reported results based on linkage of children's social care records (by virtue of the study inclusion criteria), but challenges were identified, such as data quality issues affecting the matching process. This included missing information in local authority data-sets and the complex, time-consuming nature of the process. Missing data are common in administrative data-sets and although there is a range of principled statistical methods for imputing missing values (Perkins, 2018), in some cases, the extent of missing values means that information cannot reliably be used. Doidge and Harron (2019) suggest that linkage errors, such as missed links between records that relate to the same person, or false links between unrelated records, can ultimately lead to information and selection bias. The wider challenges, limitations and technical details in data linkage research are further discussed by Lyons *et al.* (2009), Doidge and Harron (2019) and Harron *et al.* (2017).

Another caveat in using data linkage for children's social care research is the variation in data quality and recording across local authorities. Clark *et al.* (2017) found that the completeness of records in their Scottish study of looked after children and dental health needs varied by local authority. For example, a minority of LAs did not have valid SCN recorded in the CLAS data, despite SCNs being imperative to the linkage process. This could be due to the uniqueness of the care population; it is well known that the care system can often be a fractured experience for children and young people because of multiple placements, school changes and moves between local authorities. This has implications for

the data linkage process and accurately linking records across different sources.

Typically, local authority administrative data from children's social care have not been easily accessible, with limited governance of its use for research purposes (Lee *et al.*, 2022). However, this field is rapidly advancing and it is becoming more accessible; for example, the SAIL Databank (Jones, 2019) now holds population-level children's social care data for Wales which can be linked to other administrative data. In England, the Economic and Social Research Council are supporting the Education and Child Health Insights from Linked Data project, led by UCL (McGrath-Lone, 2021). As such, researchers are now in a position to not only use stand-alone administrative social care data, but also safely and securely linked to other data sources to answer research questions. When reporting studies of this nature, investigators should utilise the 'Reporting of Studies Conducted using Observational Routinely-collected Data' (RECORD) standards (RECORD-Statement, 2019), to ensure transparency and consistency in reporting methods and results. Guidance for Information about Linking Data-sets (GUILD) is another supportive tool for data linkers, analysts and researchers to access information about access, requesting and analysing linked data (Gilbert *et al.*, 2018). Whilst not a protocol or checklist, it provides guidance on managing and assessing linkage error and its impact on results.

Alongside data linkage infrastructure, mutual support and advice are available for researchers who are considering using children's social care administrative data, via the data users group set up by researchers from Oxford and UCL (see https://cscdug.co.uk/) as well as an affiliated group in Wales. Such initiatives are important for building research capacity in this area. There is a need for capacity building amongst social work researchers in particular, as social work is a field where in the UK, quantitative research capacity is relatively underdeveloped (Scourfield *et al.*, 2018). Currently, a lot of data linkage research is undertaken by medical research teams. Interdisciplinary collaboration is vital between researchers with data linkage experience and those who know the social care field intimately, through research, practice or lived experience.

Strengths and limitations

This was a rapid and not a full PRISMA systematic review; however, extensive searches were conducted of key databases, reference lists and some grey literature. We acknowledge the limitation of our search terms including the names of the national data sets and that some studies using local administrative data may not have used the national data set titles, even if using the same data. Despite this we believe the review's novelty is a strength. As far as we are aware there are no other reviews

synthesising studies that have linked children's social care data to various data sources. It is hoped that the review encouragers researchers to utilise available data that can answer complex children's social care issues within interdisciplinary study teams.

Conclusion

Children's social care administrative data are a rich source of information that can be used for research purposes. As this review shows, such data can also be linked to other sources to provide a holistic picture of issues relating to the health, education and wellbeing of children and families receiving social care support. This review provides information about the challenges and benefits for researchers when using linked routinely collected children's social care administrative data. It is hoped that with increased access to the data, feedback on the quality, scope and potential changes to local authorities' data capture with marginal cost implications could improve the utility of these valuable administrative data assets. The review has highlighted that most researchers matched individual children's social care records to other data sources but applied caution to inferring results because of missing and inaccurate records. Nevertheless, this review provides an up to date overview of the evidence on data linkage in children's social care with the hope of encouraging more researchers in the area to consider using this valuable resource.

Funding

Conflict of interest statement. None declared.

References

Baldwin, H., Biehal, N., Cusworth, L., Wade, J., Allgar, V. and Vostanis, P. (2019) 'Disentangling the effect of out-of-home care on child mental health', *Child Abuse & Neglect*, **88**, pp. 189–200.

Baldwin, H., Biehal, N., Allgar, V., Cusworth, L. and Pickett, K. (2020) 'Antenatal risk factors for child maltreatment: Linkage of data from a birth cohort study to child welfare records', *Child Abuse and Neglect*, **107**, pp. 1–11.

Berridge, D., Luke, N., Sebba, J., Strand, S., Cartwright, M., Staples, E., Mcgrath-Lone, L., Ward, J. and O'higgins, A. (2020) *Children in Need and Children in Care: Educational Attainment and Progress*, Oxford, Rees Centre.

Brownell, M. D. and Jutte, D. P. (2013) 'Administrative data linkage as a tool for child maltreatment research', *Child Abuse & Neglect*, **37**(2–3), pp. 120–4.

Bywaters, P., Scourfield, J., Jones, C., Sparks, T., Elliott, M., Hooper, J., McCartan, C., Shapira, M., Bunting, L. and Daniel, B. (2020) 'Child welfare inequalities in the four nations of the UK', *Journal of Social Work*, **20**(2), pp. 193–215.

Bywaters, P., Brady, G., Bunting, L., Daniel, B., Featherstone, B., Jones, C., Morris, K., Scourfield, J., Sparks, T. and Webb, C. (2018) 'Inequalities in English child protection practice under austerity: A universal challenge?', *Child & Family Social Work*, **23**(1), pp. 53–61.

Children's Commissioner. (2017, 2018, 2019, 2020) *Stability Index for Children in Care*. Available at: https://www.childrenscommissioner.gov.uk/wp-content/uploads/2020/11/cco-stability-index-2020.pdf (accessed 10 March 2022).

Clark, D., King, A., Sharpe, K., Connelly, G., Elliott, L., Macpherson, L. M. D., McMahon, A. D., Milligan, I., Wilson, P., Conway, D. I. and Wood, R. (2017) 'Linking routinely collected social work, education and health data to enable monitoring of the health and health care of school-aged children in state care ("looked after children") in Scotland: a national demonstration project', *Public Health*, **150**, pp. 101–11.

Cusworth, L., Bedston, S., Alrough, B., Broadhurst, K., Johnson, R., Akbari, A. and Griffiths, L. (2021) *Uncovering Private Family Law: Who's Coming to Court in England*, London, Nuffield Family Justice Observatory.

Dickens, J., Masson, J., Garside, L., Young, J. and Bader, K. (2019) 'Courts, care proceedings and outcomes uncertainty: The challenges of achieving and assessing "good outcomes" for children after child protection proceedings', *Child & Family Social Work*, **24**(4), pp. 574–81.

Doidge, J. C. and Harron, K. L. (2019) 'Reflections on modern methods: Linkage error bias', *International Journal of Epidemiology*, **48**(6), pp. 2050–60.

Edwards, R., Hughes, C. and Williams, M. (2015) *Data Linkage: Ethical and Social Concerns*. Available at: https://www.ncrm.ac.uk/news/show.php?article=5444 (accessed 10 March 2022).

Elliott, M. (2020) 'Child Welfare Inequalities in a Time of Rising Numbers of Children Entering Out-of-Home Care', *The British Journal of Social Work*, **50**(2), pp. 581–97.

Emmott, E., Jay, M. and Woodman, J. (2019) 'Cohort Profile: Children in Need Census (CIN) Records of Children Referred for Social Care Support in England', *BMJ Open*, **9**(2), pp. e023771.

Gilbert, R., Lafferty, R., Hagger-Johnson, G., Harron, K., Zhang, L., Smith, P., Dibben, C. and Goldstein, H. (2018) 'GUILD: Guidance for information about linking datasets', *Journal of Public Health (Oxford, England)*, **40**(1), pp. 191–8.

Harron, K., Dibben, C., Boyd, J., Hjern, A., Azimaee, M., Barreto, M. and Goldstein, H. (2017) 'Challenges in administrative data linkage for research', *Big Data and Society*, 1–12.

Hooper, J., Cusworth, L. and Whincup, H. (2019) *Permanently Progressing? Building Secure Futures for Children in Scotland, Linking Two Administrative Datasets about Looked after Children: Testing Feasibility and Enhancing Understanding*, Scotland, University of Stirling.

Hurren, E., Stewart, A. and Dennison, S. (2017) 'New methods to address old challenges: The use of administrative data for longitudinal replication studies of child maltreatment', *International Journal of Environmental Research and Public Health*, **14**(9), pp. 1066.

Jay, M. and Gilbert, R. (2021) 'Special educational needs, social care and health', *Archives of Disease in Childhood*, **106**(1), pp. 83–5.

Jay, M., Woodman, J., Broadhurst, K. and Gilbert, R. (2017) *Who Cares for Children? Population Data for Family Justice Research*. Available at http://wp.lancs.ac.uk/observatory-scoping-study/files/2017/10/FJO-NATIONAL-DATA-SCOPING-FINAL.pdf (accessed 10 March 2022).

Johnson, R., Ford, D., Broadhurst, K., Cusworth, L., Jones, K., Akbari, A., Bedston, S., Alrouh, B., Doebler, S., Lee, A., Smart, J., Thompson, S., Trinder, L. and Griffiths, L. (2020) 'Data Resource: Population level family justice administrative data with opportunities for data linkage', *International Journal of Population Data Science*, **5**(1), pp. 1–14.

Jolley, R. J., Lorenzetti, D. L., Manalili, K., Lu, M., Quan, H. and Santana, M. J. (2017) 'Protocol for a scoping review study to identify and classify patient-centred quality indicators', *BMJ Open*, **7**(1), pp. e013632.

Jones, K. H., Ford, D. V., Thompson, S. and Lyons, R. A. (2019) 'A profile of the SAIL databank on the UK secure research platform', *International Journal of Population Data Science*, **4**(2), pp. 1–12.

Lee, A., Elliott, M., Scourfield, J., Bedston, S., Broadhurst, K., Ford, D. and Griffiths, L. (2022) 'Data Resource: Children Receiving Care and Support and Children in Need, Adminstrative Records in Wales', *International Journal of Population Data Science*. (in press).

Lugg-Widger, F., Cannings-John, R., Angel, L., Moody, G., Segrott, J., Kenkre, J. and Robling, M. (2018) 'Assessing the impact of specialist home visiting upon maltreatment in England: A feasibility study of data linkage from a public health trial to routine health and social care data', *Pilot and Feasibility Studies*, **4**(98), pp. 98–11.

Lyons, R., Jones, K., John, K., John, G., Brooks, C., Verplancke, P., Ford, D., Brown, G. and Leake, K. (2009) 'The SAIL databank: Linking multiple health and social care datasets', *BMC Medical Informatics and Decision Making*, **9**(1), pp. 1–8.

McCartan, C., Bunting, L., Bywaters, P., Davidson, G., Elliott, M. and Hooper, J. (2018) 'A four-nation comparison of kinship care in the UK: The relationship between formal kinship care and deprivation', *Social Policy and Society*, **17**(4), pp. 619–35.

McGrath-Lone, L., Blackburn, R. and Gilbert, R. (2021) *The Education and Child Health Insights from Linked Data (ECHILD) Database: An Introductory Guide for Researchers*. Available at: https://www.ucl.ac.uk/child-health/sites/child-health/files/echild_user_guide_v1.1.0.pdf (accessed 10 March 2022).

McGrath-Lone, L., Harron, K., Dearden, L., Nasim, B. and Gilbert, R. (2016) 'Data resource profile: Children looked after return (CLA)', *International Journal of Epidemiology*, **45**(3), pp. 716–7.

McMahon, A. D., Elliott, L., Macpherson, L. M. D., Sharpe, K. H., Connelly, G., Milligan, I., Wilson, P., Clark, D., King, A., Wood, R. and Conway, D. I. (2018) 'Inequalities in the dental health needs and access to dental services among looked

after children in Scotland: A population data linkage study', *Archives of Disease in Childhood*, **103**(1), pp. 39–43.

Pearson, R. J., Jay, M. A., Wijlaars, L. P. M. M., De Stavola, B., Syed, S., Bedston, S. J. and Gilbert, R. (2020) 'Association between health indicators of maternal adversity and the rate of infant entry to local authority care in England: A longitudinal ecological study', *BMJ Open*, **10**(8), pp. e036564.

Perkins, N. J., Cole, S. R., Harel, O., Tchetgen Tchetgen, E. J., Sun, B., Mitchell, E. M. and Schisterman, E. F. (2018) 'Principled approaches to missing data in epidemiologic studies', *American Journal of Epidemiology*, **187**(3), pp. 568–75.

PRISMA. (2020) *PRISMA 2020 Checklist*. Available at: http://prisma-statement.org/ (accessed 27 September 2021).

RECORD-Statement. (2019) *Record-Statement*. Available at: https://www.record-statement.org/ (accessed 24 May 2021).

Scourfield, J., Shardlow, S., Rees, A. and Zhang, M. (2018) 'A profile of UK doctoral candidates in social work and social care', *The British Journal of Social Work*, **48**(8), pp. 2313–31.

Sebba, J., Berridge, D., Luke, N., Fletcher, J., Bell, K., Strand, S., Thomas, S., Sinclair, I. and O'Higgins, A. (2015) *The Educational Progress of Looked After Children in England: Linking Care and Educational Data*, Oxford, Rees Centre.

Sinclair, I., Luke, N., Fletcher, J., O'Higgins, A., Strand, S., Berridge, D., Sebba, J. and Thomas, S. (2020) 'The education of children in care and children in need: Who falls behind and when?', *Child & Family Social Work*, **25**(3), pp. 536–47.

Sinclair, I., Luke, N. and Berridge, D. (2019) 'Children in care or in need: Educational progress at home and in care', *Oxford Review of Education*, **45**(4), pp. 443–60.

Teyhan, A., Boyd, A., Wijedasa, D. and MacLeod, J. (2019) 'Early life adversity, contact with children's social care services and educational outcomes at age 16 years: UK birth cohort study with linkage to national administrative records', *BMJ Open*, **9**(10), pp. e030213.

Webb, C., Bywaters, P., Scourfield, J., Davidson, G. and Bunting, L. (2020a) 'Cuts both ways: Ethnicity, poverty, and the social gradient in child welfare interventions', *Children and Youth Services Review*, **117**, pp. 105299–13.

Webb, C., Bywaters, P., Scourfield, J., McCartan, C., Bunting, L., Davidson, G. and Morris, K. (2020b) 'Untangling child welfare inequalities and the "Inverse Intervention Law" in England', *Children and Youth Services Review*, **111**, pp. 104849–10.

Woollard, M. (2014) 'Administrative data: Problems and benefits. A perspective from the United Kingdom', in Dusa, A. (ed.), *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences*, Berlin, SCIVERO.

Zhang, M. L., Boyd, A., Cheung, S. Y., Sharland, E. and Scourfield, J. (2020) 'Social work contact in a UK cohort study: Under-reporting, predictors of contact and the emotional and behavioural problems of children', *Children and Youth Services Review*, **115**, pp. 105071–9.