
Publishers page: https://doi.org/10.1093/bjsw/bcac049 

Please note: Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
Use of Linked Administrative Children’s Social Care Data for Research: A Scoping Review of Existing UK Studies

G. Allnatt¹, M. Elliott², J. Scourfield², A. Lee¹ and L. J. Griffiths¹,*

¹Population Data Science, Swansea University Medical School, Swansea SA2 8PP, UK
²CASCADE, 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

Accepted: February 2022
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.
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
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>
<thead>
<tr>
<th>Study</th>
<th>Output type</th>
<th>Sample size and age</th>
<th>National or local level study</th>
<th>Country</th>
<th>Data-sets linked</th>
<th>Data-set type</th>
<th>Linkage process</th>
<th>Main results and conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Studies covering whole UK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bywaters et al. (2020)</td>
<td>Published research article</td>
<td>( 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)</td>
<td>Sub-national</td>
<td>UK</td>
<td>CLA; CP register and plans</td>
<td>Individual-level social care data</td>
<td>Deprivation indices; child population measures.</td>
<td>Area- and neighbourhood-level administrative data.</td>
</tr>
<tr>
<td>McCartan et al. (2018)</td>
<td>Published research article</td>
<td>( N = 55 ) LAs; twenty-two from Wales, eighteen from England, eleven from N.I. and ten from Scotland. (age 0–17)</td>
<td>Sub-national</td>
<td>UK</td>
<td>CLA; CP plan</td>
<td>Individual-level social care data</td>
<td>Deprivation indices; child population measures.</td>
<td>Area- and neighbourhood-level administrative data.</td>
</tr>
<tr>
<td><strong>Studies covering only England</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jay and Gilbert (2021)</td>
<td>Published research article</td>
<td>( N = 475,363 ) (age 5–16)</td>
<td>National</td>
<td>England</td>
<td>CIN; CLA NPD</td>
<td>Individual-level social care data.</td>
<td>Individual-level school registration and attainment administrative data.</td>
<td>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.</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Output type</th>
<th>Sample size and age</th>
<th>National or local level study</th>
<th>Country</th>
<th>Data-set information</th>
<th>Main results and conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baldwin et al. (2020)</td>
<td>Published research article</td>
<td>N = 11,332 (age 50–100 months)</td>
<td>Local</td>
<td>England</td>
<td>Individual-level social care data.</td>
<td>Maternal smoking, mental ill health, lower maternal education level and other indicators of deprivation are salient risk factors for child maltreatment.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Individual-level questionnaire data.</td>
<td>Placement and educational instability linked with poorer attainment. Need for better visibility of the CIN group and the importance of early intervention.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Individual-level school registration and attainment administrative data.</td>
<td>Placement and educational instability linked with poorer attainment. Need for better visibility of the CIN group and the importance of early intervention.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Individual-level child-level social worker history.</td>
<td>Placement and educational instability linked with poorer attainment. Need for better visibility of the CIN group and the importance of early intervention.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Individual-level school registration and attainment administrative data.</td>
<td>Placement and educational instability linked with poorer attainment. Need for better visibility of the CIN group and the importance of early intervention.</td>
</tr>
<tr>
<td>Study</td>
<td>Output type</td>
<td>Sample size and age</td>
<td>National or local level study</td>
<td>Country</td>
<td>Data-set information</td>
<td>Linkage process</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------------------</td>
<td>---------------------</td>
<td>-------------------------------</td>
<td>---------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Pearson et al. (2020)</td>
<td>Published research article</td>
<td>N = 131 Las</td>
<td>National</td>
<td>England</td>
<td>CLA, Individual-level social care data. Individual-level health data.</td>
<td>Aggregated values per local authority in both data-sets were linked according to the name of the LA.</td>
</tr>
<tr>
<td>Sinclair et al. (2020)*</td>
<td>Published research article</td>
<td>N = 642,805 (age 0–16)</td>
<td>National</td>
<td>England</td>
<td>CLA, Individual-level social care data. Individual-level school registration and attainment administrative data.</td>
<td>Pseudonymised UPD used to link the CLA records to the NPD.</td>
</tr>
<tr>
<td>Webb et al. (2020a)</td>
<td>Published research article</td>
<td>N = 52,179 (age 0–17)</td>
<td>National</td>
<td>England</td>
<td>CIN; CLA; CP Plan, LSOA code in which children lived, or from where they entered care, used to link social care data to LSOA-level administrative data.</td>
<td>LA-level aggregated administrative data.</td>
</tr>
<tr>
<td>Webb et al. (2020b)</td>
<td>Published research article</td>
<td>N = 860,000 (age 0–17)</td>
<td>National</td>
<td>England</td>
<td>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</td>
<td>MSOA code in which children lived, or from where they entered care, used to link social care data to MSOA-level administrative data.</td>
</tr>
<tr>
<td>Study</td>
<td>Output type</td>
<td>Sample size and age</td>
<td>National or local level study</td>
<td>Country</td>
<td>Data-sets linked</td>
<td>Data-set type</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>---------------------</td>
<td>-------------------------------</td>
<td>---------</td>
<td>-----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Zhang et al. (2020)</td>
<td>Published research article</td>
<td>$N = 14,062$ (age 21–73 months)</td>
<td>Local</td>
<td>England</td>
<td>CP Register</td>
<td>Individual-level social care data.</td>
</tr>
<tr>
<td>Sinclair et al. (2019)</td>
<td>Published research article</td>
<td>$N = 642,805$ (main cohort) $N = 6,236$ (CLA) $N = 20,383$ (CIN)</td>
<td>National</td>
<td>England</td>
<td>CIN; CLA</td>
<td>Individual-level social care data</td>
</tr>
<tr>
<td>Baldwin et al. (2019)</td>
<td>Published research article</td>
<td>$N = 363$ (age 0–7)</td>
<td>Local</td>
<td>England</td>
<td>Child welfare data-set Survey of social workers; interviews with children’s caregivers</td>
<td>Individual-level social care data</td>
</tr>
<tr>
<td>Teyhan et al. (2019)</td>
<td>Published research article</td>
<td>$N = 12,868$ (age 0–18)</td>
<td>Local</td>
<td>England</td>
<td>CIN; CLA</td>
<td>Individual-level social care data</td>
</tr>
</tbody>
</table>
Table 2. (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Output type</th>
<th>Sample size and age</th>
<th>National or local level study</th>
<th>Country</th>
<th>Data-set information</th>
<th>Main results and conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lugg-Widger et al. (2018)</td>
<td>Published research article</td>
<td>$N = 1,545$ (aged under 6 years old)</td>
<td>Local England</td>
<td>CLA; CIN</td>
<td>Individual-level social care 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.</td>
<td>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.</td>
</tr>
<tr>
<td>Bywaters et al. (2018)</td>
<td>Published research article</td>
<td>$N = 1,432,180$ (age 0–17)</td>
<td>National England</td>
<td>CLA; CIN</td>
<td>Individual-level social care data linked to LSOA- and local authority-level administrative data.</td>
<td>Relationship identified between deprivation and intervention rates and significant inequalities between ethnic categories.</td>
</tr>
<tr>
<td>Sebba et al. (2015)</td>
<td>Report</td>
<td>$N = 622,970$ (aged 15 years on 12 September 2012)</td>
<td>National England</td>
<td>CLA; NPD</td>
<td>Individual-level social care data linked to the CLA records to the NPD.</td>
<td>CIN did least well academically. SEN common amongst the CLA population.</td>
</tr>
<tr>
<td>Study</td>
<td>Output type</td>
<td>Sample size and age</td>
<td>National or local level study</td>
<td>Country</td>
<td>Data-set information</td>
<td>Data-sets linked</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>---------------------</td>
<td>-------------------------------</td>
<td>---------</td>
<td>----------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Studies covering England and Wales</td>
<td>Dickens et al. (2019)</td>
<td>Published research article</td>
<td>$N = 616$ (age 0–17)</td>
<td>Local</td>
<td>England and Wales (sample from six LAs in England and Wales)</td>
<td>CIN, CLA (England and Wales) Social work case files* For a subsample of children from Cafcass database</td>
</tr>
<tr>
<td>Studies covering only Wales</td>
<td>Elliott (2020)</td>
<td>Published research article</td>
<td>$N = 8,853$ (age 0–17)</td>
<td>National</td>
<td>Wales</td>
<td>CLA data Child population estimates; Welsh index of multiple deprivation (WIMD) LSOA from where children entered care, used to link social care data to administrative data.</td>
</tr>
<tr>
<td>Studies covering only Scotland</td>
<td>Hooper et al. (2019)</td>
<td>Report</td>
<td>$N = 1,834$ (CLAS) $N = 1,396$ (SCRA) (age 0–5)</td>
<td>National</td>
<td>Scotland</td>
<td>CLA Statistics (Scotland) Scottish Children’s Reporter Administration</td>
</tr>
</tbody>
</table>
Table 2. (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Output type</th>
<th>Sample size and age</th>
<th>National or local level study</th>
<th>Country</th>
<th>Data-sets linked</th>
<th>Data-set type</th>
<th>Linkage process</th>
<th>Main results and conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>McMahon et al. (2018)</td>
<td>Published research article</td>
<td>N = 10,924 (age 5–11)</td>
<td>National</td>
<td>Scotland</td>
<td>CLA statistics (Scotland)</td>
<td>Individual-level social care data</td>
<td>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.</td>
<td>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.</td>
</tr>
<tr>
<td>Clark et al. (2017)</td>
<td>Published research article</td>
<td>N = 637,815 (age 4–19)</td>
<td>National</td>
<td>Scotland</td>
<td>CLA statistics (Scotland)</td>
<td>Individual-level social care data</td>
<td>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.</td>
<td>Variation in completeness of individual records across local authorities which impacts on linkage. CLA have poorer health outcomes than their non-looked after counterparts.</td>
</tr>
</tbody>
</table>

*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 Commissioner, 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 datasets 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 encourages 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

The Children’s Social Care Research and Developmental Centre (CASCADE) partnership receives infrastructure funding from Health and Care Research Wales (HCRW) (Grant number-517199).

Conflict of interest statement. None declared.

References


