Data Explained

—

Scottish Government's Looked After Children Longitudinal Dataset

Authors: Joanna Soraghan, University of Strathclyde and Gillian Raab, University of Edinburgh
Date: April 2023

Introduction

The Scottish Government Looked After Children longitudinal dataset (hereafter referred to as the longitudinal CLAS dataset) contains person-level records for all children and young people who have been placed under the care of a local authority in Scotland from 2008 onwards. It contains key information relating to demographics, placement settings, and legal reasons for being in care.

This Data Explained summarises experiences and learning from working with the longitudinal CLAS dataset in the course of producing research into the impact of the Covid-19 pandemic on children's care journeys in Scotland. It provides an outline of the CLAS dataset, summarising the variables available, issues within the dataset, and current processes for data access. It is hoped that this information will be useful in guiding future researchers in the use of this data and in providing feedback to aid future dataset development and documentation.

The data discussed in this Data Explained comes from the Scottish Government and was accessed through the National Safe Haven, a secure research environment. All data was anonymised prior to being shared with researchers.

How is the data collected?

The dataset is comprised of annual returns sent from local authorities to Scottish Government from the reporting year 2008/09 onwards (with 2020/21 being the most recent year available at the time of writing)[1]. These returns contain information on all children and young people that have been under the care of the local authority at any point within the previous reporting year. Some historic data from earlier time periods is also included where a child was in care before 1 April 2008 and continued to be in care beyond this date. These individual annual returns have been compiled in order to create the longitudinal CLAS dataset. Full details on how these returns were collated and issues encountered in the process have been provided by the Scottish Government in their '[Looked after children in Scotland - longitudinal data user guide](#)'.

The current iteration of the dataset (covering the period to July 2021) includes information on almost 65,000 individuals, with details on approximately:

- 76,000 episodes of care
- 162,000 placements
- 203,000 legal reasons.

A huge benefit of having the data in this longitudinal format is that the records of each individual child are linked over time. This provides opportunities for more complex, longitudinal analysis to gain insight into the journeys of children and young people throughout their time in care.

---

[1] As of 2010/11, the reporting year for this data runs from 1 August to 31 July of the following year. It is useful to note, however, that data collection for the period 2008-09 was from 1 April 2008 to 31 March 2009, and for 2009-10 was from 1 April 2009 to 31 July 2010 (i.e. a longer 15-month period).

Note: While a standalone version of the dataset containing data to July 2021 was made available to the researchers for the purposes of their project, the version that is currently ready and available for linkage runs only to July 2019. A more up-to-date version for linkage (including data to July 2022) is expected to be available in the coming months.

Key variables

Information available within the longitudinal CLAS dataset is divided into 'Placement' data and 'Legal Reasons' data, and includes:

- ❖ details of the child in care (e.g., date of birth, gender, ethnicity, disability status)
- ❖ the dates that the individual has been in care
- ❖ the local authority within which they were looked after
- ❖ the legal basis (or bases) by which they were in care
- ❖ the types of placement settings they have been accommodated in (e.g., foster care, with friends and relatives, looked after at home with their parents, residential care)
- ❖ destinations upon leaving care (e.g., home with parents, adopted, with friends and relatives, own tenancy)
- ❖ whether the child has a pathway plan in place if they are leaving care over the age of 16 years
- ❖ the reporting year in which the information was returned to the Scottish Government (the 'extract' year).

Data in both the 'Placements' and 'Legal Reasons' tables are clustered within episodes, where an 'episode' of care refers to a period of time during which a child is continuously under the care of the local authority. Each episode can be comprised of more than one placement of different types and a child should have a legal reason in place at any time they are in care, however the dates for placements and legal reasons do not necessarily align.

A full list of available variables, along with further information on the variables, can be found within the provided metadata for this dataset which is available in the [ADR Scotland Data Catalogue](#).

Note: While carrying out data cleaning, the authors found that the 'extract' variable – an indicator of the reporting year in which a given record was returned to the Scottish Government - was extremely useful. While it will likely not be related to the research question of interest, it allows users to check for potentially outdated records and know which information is most recent (and likely most reliable) where there are discrepancies in the data. For this reason, we would advise that this variable be requested when applying for access to the data.

What can the data be used for?

The [Administrative Data Research (ADR) Scotland](#) programme supports making data available and more readily linkable for research purposes in the public benefit. An aim of the programme is to help make administrative datasets more readily linkable in order to provide richer research opportunities.

The ADR Scotland Data Catalogue provides an idea of other datasets (e.g., Pupil Census, Qualifications, Child Protection) that the longitudinal CLAS dataset could potentially be linked to, however this list is not exhaustive. For example, work is currently underway to link the dataset to both health visiting data from Public Health Scotland and data on children's hearings from the Scottish Children's Reporter Administration for an upcoming Scottish Centre for Administrative Research (SCADR) research project titled 'Growing Up In Kinship Care'.

Access to the longitudinal CLAS data as a standalone dataset is gained through submitting an application form to the Scottish Government. Prior to completing the form, it is strongly advisable to contact the Scottish Government's Children and Families Analysis team to discuss your planned work and data requirements. Contact details can be found on the Scottish Government website.

For data linkage projects, the electronic Data Research and Innovation Service (eDRIS) team at Public Health Scotland should be contacted in the first instance. eDRIS provides support to researchers who wish to make use of administrative datasets by providing a single-entry point and end-to-end support to researchers throughout the process - from information governance approvals to data linkage and access. More information on this service is available on the eDRIS website. The authors advise that the data linkage process can be lengthy, and it is worth having discussions around this and factoring timescales into project proposals and planning from an early stage.

Existing research or examples of previous research

The longitudinal CLAS dataset has already been utilised for research in a variety of ways.

❖ The Children's Health in Care in Scotland (CHiCS) study (Allik et al, 2021) linked the longitudinal CLAS dataset to health data. This data linkage enabled researchers to provide population-wide evidence on the health outcomes of children with care experience, across a variety of outcomes such as hospitalisations, prescriptions, pregnancies, and mortality. Read more on the study and its findings.

❖ McMahon et al (2018) linked the CLAS dataset to the Pupil Census and NHS dental datasets to assess inequalities in dental health between looked-after children and those who had not experienced care.

❖ The dataset has also been used to facilitate an analysis of the rates at which infants under the age of one have been coming into care in Scotland in recent years (Raab et al, 2020). The longitudinal nature of the data allowed for the care journey of these infants to be explored in a way that would not have been possible with annual 'snapshot' data. Findings from the 'Infants Born into Care in Scotland' 2020 study were based on data up until 2017 and can be found at SCADR's website. You can also read an update to this report that includes data from 2021, published in 2023.

- ❖ The upcoming SCADR '**Growing Up in Kinship Care**' project was mentioned earlier in this guide. This project will look to link the longitudinal CLAS dataset to data from health, education, child protection and the children's hearing system. This will allow researchers to gain a richer picture of what is currently happening for children living in kinship care, where a formal decision has been made that a child should live with a relative separate from their parent(s).

- ❖ The learning shared in this Data Explained was gained while conducting research into **The impact of the Covid-19 pandemic on children's care journeys in Scotland**. You can access **findings from this research**.

Data limitations encountered

The data was not originally collected for research and it is expected that there are gaps and inconsistencies in its recording, as with all administrative data.

As is detailed in the [Looked after children in Scotland - Longitudinal Data User Guide](#), the dataset comes with built in data quality flags that have been added by the Scottish Government in the process of creating the longitudinal dataset. These highlight where there are inconsistencies in the data or where records do not meet a requirement of the data. For example, this could include instances where a placement is reported to have started prior to the child's date of birth, or where the child is recorded as having two overlapping episodes of care. The guide contains further details on the data quality flags (including their prevalence), alongside suggestions for how the flagged data quality issues can be addressed.

Depending on the research questions and intended analysis, time will likely need to be built into the research project for data cleaning. Information on the data cleaning process implemented by the authors for the current work can be found in the appendix to this document.

There have been changes to the way in which disability information is recorded over the period from 2008-2021, and these may need to be taken into account when conducting analysis using this variable. Further details on these changes can be found in the metadata document in the ADR Scotland Data Catalogue.

Data quality was generally found to be better for placements as opposed to legal reasons data. Further details on this are given in the appendix.

Recommendations to data owners

Recent changes in legislation[2] mean that local authorities now have a duty to ensure siblings are supported to stay together, where appropriate. Unfortunately, the administrative data that is currently available does not provide a means for identifying sibling groups or analysing how effectively this legislation has been implemented. The authors understand, however, that amendments will be made to the data returns from 2023/24 in order to provide better information on whether children are in placements with their siblings, and we welcome these amendments.

---

[2] Part 13 of the Children (Scotland) Act 2020 and the Looked After Children (Scotland) Amendment Regulations 2021.

When data is returned to Scottish Government by the individual local authorities, it is checked against previous years and any issues or inconsistencies are queried at that stage. The authors believe that there could be value in the provision of a tool for local authorities for interim data quality checking, not just end of year returns – similar to the **validation tool** provided by **Data to Insight** for local authorities in England. Access to more regular verification could allow local authorities to identify issues sooner and potentially have more success in rectifying these prior to returning data to government.

Finally, certain variables within the dataset are comprised of many categories, some of which have very low counts on a monthly/annual basis. This can be problematic for both publication of raw figures and ensuring that numbers are feasible for analysis. The authors could see value in the provision of derived, simplified versions of certain variables (such as placement types or legal reasons), or suggestions from the data owners on how categories should best be combined where necessary.

Conclusion

In recent years, datasets such as these have become more widely available for the purposes of research in the public benefit. This development has the potential to drastically improve the evidence base around which crucial policy and budgetary decisions are made. It is essential that there are strong ethical guidelines around the use of data. The authors welcome the transparency and opportunities that this wider access and greater insight can bring and feel that this is an important step towards a society that works better for us all.

References

Allik, M., Brown, D., Taylor Browne Lūka, C., Macintyre, C., Leyland, A. H., & Henderson, M. (2021). Cohort profile: The 'Children's Health in Care in Scotland' (CHiCS) study-a longitudinal dataset to compare health outcomes for care experienced children and general population children. *BMJ open*, *11*(9), e054664. DOI: **https://doi.org/10.1136/bmjopen-2021-054664**

McMahon AD, Elliott L, Macpherson LM, et al (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 2018;**103**:39-43 DOI: **https://doi.org/10.1136/archdischild-2016-312389**

Raab G, MacIntyre C & McGhee J (2020). Infants born into care in Scotland: Initial findings. Scottish Centre for Administrative Data Research (SCADR). DOI: **https://doi.org/10.7488/era/732**

Soraghan J, Raab G & Troncoso P (2023), The Impact of Covid-19 on Children's Care Journeys in Scotland: An Analysis of the Administrative Data on Looked After Children (SCADR). DOI: **http://dx.doi.org/10.7488/era/3245**

Raab G, Soraghan J, MacIntyre C, McGhee J & Troncoso P (2023). Infants born into care in Scotland: Updated findings. Scottish Centre for Administrative Data Research (SCADR). (Coming April 2023).

Glossary

Term	Meaning
Looked-after children	Children who are in the care of their local authority.
Placement	The place in which a looked after child is living. This can include foster care, residential care, with family or friends or at home with their parents.
Legal Reason	The legal basis by which a child is in care. This can be on a voluntary basis with cooperation from the parent(s), or via compulsory measures mandated by a court or Children's Hearings[3] panel.
Episode	A continuous period of time in which a child or young person is looked after. An episode can be comprised of multiple placements of different types.

---

[3] **https://www.chscotland.gov.uk/**

Disclaimer

Acknowledgements

Contact

Name: Joanna Soraghan, Data Analyst, Centre for Excellence for Children's Care and Protection (CELCIS), University of Strathclyde

Email: joanna.soraghan@strath.ac.uk

Appendix I: Data Cleaning

As discussed previously within the Data Limitations sections, there are inconsistencies found within the dataset. This is to be expected with all administrative datasets due to the nature with which they are collected. As such, the authors undertook a data cleaning process to ensure that the information contained within the dataset was coherent and that any duplicated, outdated or incorrect records would not skew the outcomes of the analysis. Steps were taken to ensure that:

- ❖ outdated records were removed (i.e. duplicate records provided in more than one return year)
- ❖ placements, legal reasons and episodes did not have missing start dates
- ❖ placements, legal reasons and episodes could not start before the child's date of birth
- ❖ episode start dates should match with the start of the first placement and legal reason in that episode
- ❖ episode end dates should match with the end of the final placement and legal reason in that episode
- ❖ episodes for a given child should not overlap
- ❖ placement within episodes should be consecutive (i.e., no gaps between them or overlapping placements)
- ❖ episode end dates were imputed for episodes that had been incorrectly left open (i.e., where it was known that the child had left care within a certain year, but no episode end date was provided).

We attempted to retain as many records as possible and only 0.2% of children had to be excluded from the analysis. An alternative approach would have been to exclude all children with any inconsistent records, but the authors felt that this would result in the loss of too much information and could potentially bias results.

Generally speaking, there were more data quality issues found within the legal reasons data than the placements data. For example, there appeared to be regional variations in practice around reporting these. The Glasgow local authority appears to largely return only one legal reason per reporting year, whereas other local authorities report all legal reasons that were in place within the year. Looking at the more than 1000 cases where the start of the first legal reason did not coincide with the start of the episode, over 90% were from Glasgow. Glasgow is the most populous local authority within Scotland, and this issue may be important to explore further or bear in mind when conducting any analysis of the legal reasons data.

All data cleaning and analysis was conducted within the R software environment for statistical computing. The authors are happy to share full details of the process implemented and the code utilised to future researchers where this would be helpful. To access this, please contact Joanna Soraghan at the contact details previously listed.