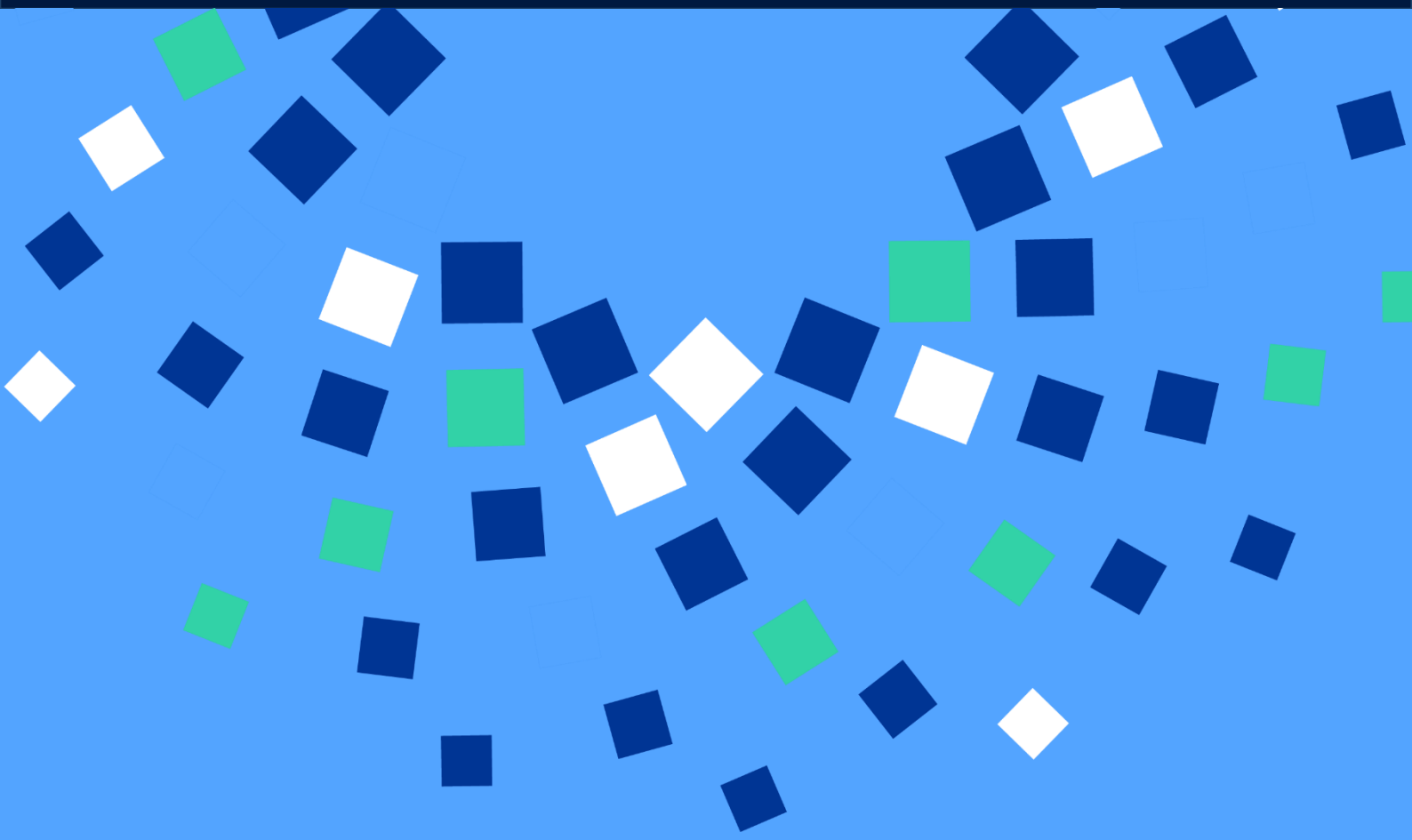
YDG CYMRU | ADR WALES

Data Explained

Annual District Birth Extract (ADBE)

Authors: Sarah Aldridge, Victoria Best, Roberta Chiovoloni, Ashley Akbari
Date: October 2026

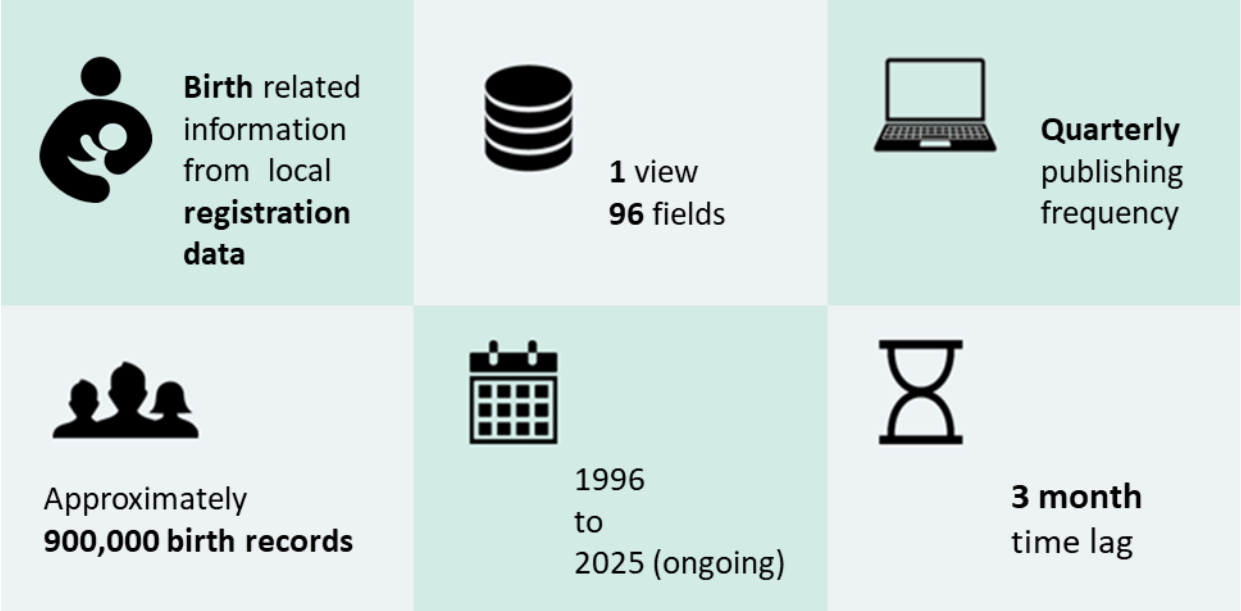


Summary

This Data Explained summarises experiences and learning from working with the Annual District Birth Extract (ADBE) data. This publication is intended to help guide future researchers using this data and to provide feedback on future data development and documentation.

The administrative data discussed in this Data Explained was made securely available through an approved SAIL project via the independent Information Governance Review Panel (IGRP) – project 1934. The data used in this research project comes from the Office for National Statistics (ONS) and was accessed through the Secure Anonymised Information Linkage (SAIL) Databank. The data were not originally collected for research, and gaps and inconsistencies in their recording are expected; some are detailed below.

Overview



Introduction

The Office for National Statistics (ONS) Annual District Birth Extract (ADBE) is a comprehensive register of all births to Welsh residents. It includes births occurring both inside and outside Wales, with subsequent registrations in Wales (Office for National Statistics 2025). Birth statistics derived from the ADBE are based on information recorded at the registration of live births and stillbirths as part of civil registration, which is a legal requirement. ADBE was established in 1996, and these data represent the most complete source of birth information available for Wales.

Local registration services carry out birth registration in partnership with the General Register Office (GRO) for England and Wales. While local registrars collect the information, ONS is the data owner and is responsible for producing official birth statistics. The provision of civil registration data from the GRO to the ONS is governed by a formal service-level agreement, ensuring consistent and secure data transfer. Since the early 2010s, the number of births to Welsh residents

has steadily declined, with recent years recording the lowest birth counts since the start of the ADBE collection (Figure 1).

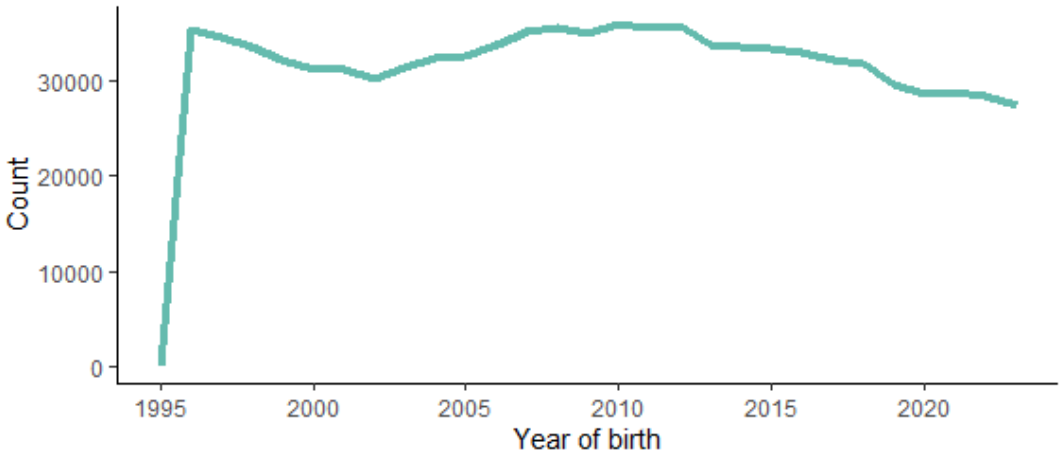


Figure 1 - The number of births recorded in ADBE per year.

How is the data collected?

In England, Wales and Northern Ireland, births must be registered at a local register office within 42 days (21 days in Scotland). Registration is usually undertaken by one or both parents, who act as the informant and provide most of the information recorded at the time of registration. These registration records are collated centrally and supplied to the ONS, the data owner. For research use in Wales, Digital Health and Care Wales (DHCW), formerly known as the NHS Wales Informatics Service (NWIS), makes the data available and provides them to the SAIL Databank in a de-identified, research-ready format.

The majority of information captured at registration is supplied directly by parents. This includes the child's date of birth (and time of birth where more than one live child is born), place of birth, sex, and name, as well as the mother's usual address. Information is also collected on each parent's country of birth, occupation, employment status, and industry, along with whether the pregnancy resulted in a multiple birth. Legislative changes introduced in 2009 allow same-sex female couples to jointly register a birth, and records from 2009 onwards reflect this (Office for National Statistics 2015).

For live births, details of birthweight are provided by the hospital or healthcare worker who attended the birth. For stillbirths, clinical information including cause of death, gestational age, and foetal weight is supplied on a medical certificate or notification completed by a healthcare provider. Causes of stillbirth are coded using the International Classification of Diseases version 10 (ICD-10) (and ICD-9 prior to 2001).

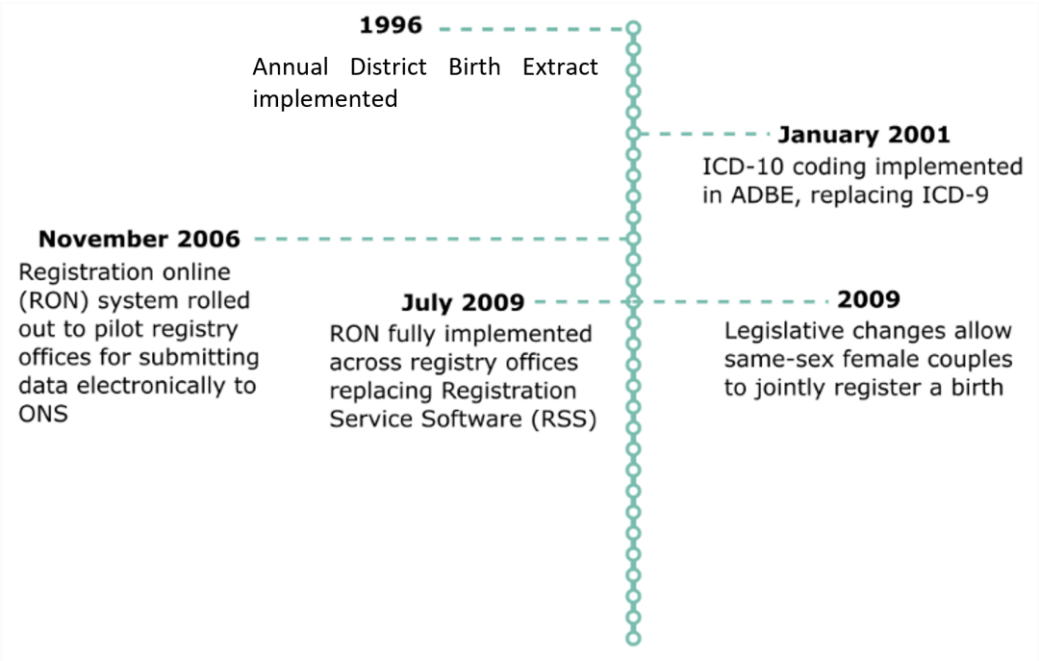


Figure 2 – Timeline of key events for ADBE.

The ADBE includes both the week of birth and the birth registration date, enabling users to assess the delay between birth occurrence and registration. The median interval between birth and registration is 18 days (Figure 2), although this has varied over time. Births registered in 1995 show longer delays, likely reflecting the small numbers during the establishment phase. Between 1996 and 2019, the median lag was relatively stable at around 17–18 days. During the COVID-19 pandemic, registration delays increased substantially, with median lags rising to 44 days in 2020 and 32 days in 2021. While delays reduced in 2022 and 2023, they remain above pre-pandemic levels and should be considered when analysing recent birth cohorts in SAIL (Figure 3).

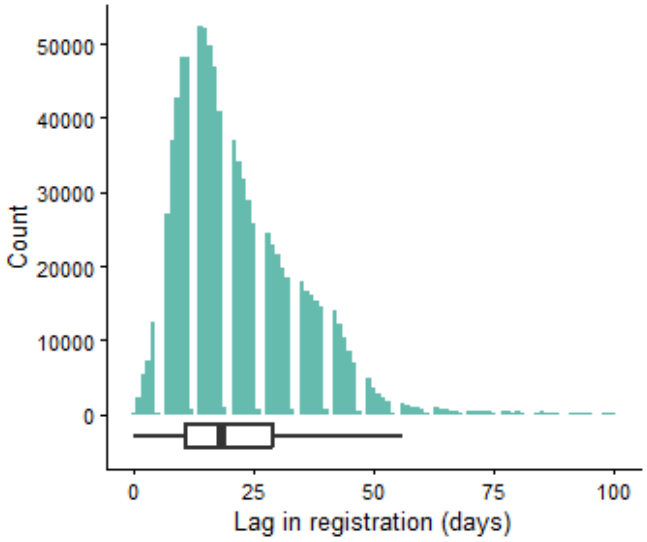


Figure 3 – The time lag of birth registration across all births in ADBE.

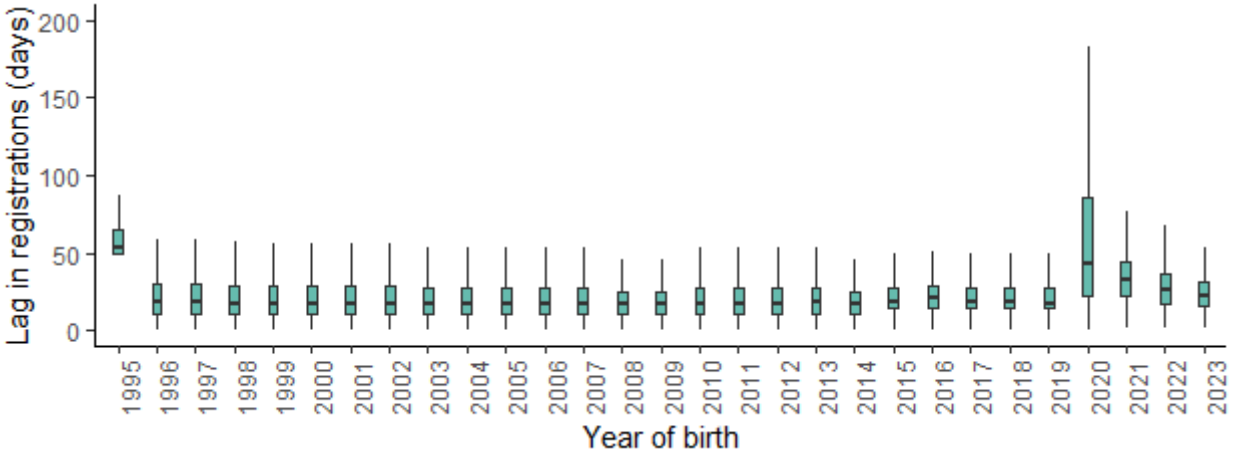


Figure 4 – The time lag for birth registration annually.

Data structure and linkage

ADBE is provided within SAIL as a single, individual-level table, with each record representing an individual live birth or stillbirth. The data includes a unique Anonymised Linkage Field (ALF) for each individual, enabling robust linkage to other data sources in the SAIL Databank.

Key variables

ADBE supports the identification of live births, stillbirths, and a range of birth-related characteristics. However, variable completeness varies substantially, reflecting differences in how information is supplied at registration and the availability of parental details.

Core variables required for linkage and cohort identification, particularly the child ALF, are highly complete. Some parental and registration-based variables also show high completeness. For example, the indicator for whether the parents were married at the time of birth (*BIRTH_MARRIAGE_IND_CD*) is complete for nearly all records. In contrast, variables describing paternal socioeconomic characteristics show much lower completion rates. Fields such as *FATHER_SOCIOECONOMIC_CLASS_CD* and *FATHER_OCC_CLASS_CD* are populated for approximately 10% of records. Researchers should account for this variability when selecting covariates and interpreting results.

For stillbirth records, data completeness is high for clinical fields. Fewer than 1% of stillbirths are missing diagnostic or cause-of-death ICD-10 codes, and up to nine additional clinical variables may be available, providing more detailed information on circumstances and outcomes.

Field	Values	Completeness
ALF_PE	Unique ID	100% coverage
ALF_STS_CD	1 or 99	99.4% are strong match quality
BIRTH_MARRIAGE_IND_CD	1,2,3,4	100% coverage
STILLBIRTH_DIAG_1_CD - *STILLBIRTH_DIAG_15_CD*	ICD-9/10 codes	<1% of births flagged as stillbirths are missing *STILLBIRTH_DIAG_1_CD* Coverage decreases with increasing *DIAG_CD, fields STILLBIRTH_DIAG_9_CD - 15_CD* are empty
FATHER_SOCIOECONOMIC_CLASS_CD	39 numeric codes	9.6% of rows contain a non-null value
FATHER_OCC_CLASS_CD	951 numeric codes	9.6% of rows contain a non-null value

Table 1 - Some key variables available in ADBE and their completeness.

What can the data be used for?

The ADBE underpins the production of official birth statistics and is used by the ONS for national reporting. Within SAIL, ADBE provides a robust foundation for research examining birth outcomes (including birthweight), fertility patterns, and parental and household characteristics across the Welsh population. Therefore, supporting longitudinal analyses that explore how early-life and birth characteristics relate to later health, social, and administrative outcomes. This makes ADBE particularly valuable for life-course research and population health studies.

For researchers focusing specifically on stillbirths, ADBE includes up to nine fields each for causes and diagnostic information, allowing for more detailed investigation of stillbirth characteristics and patterns.

Using ADBE

ADBE is a highly accurate and near-complete record of births in Wales, forming the basis of ONS reporting of Welsh birth statistics. However, as with all administrative datasets, careful consideration of data structure, coding systems, and completeness is essential prior to analysis. Users are strongly advised to consult the accompanying SAIL metadata and documentation to ensure appropriate interpretation and use of variables.

Researchers should pay particular attention to the coding frameworks used within the data, including changes in ICD versions over time for stillbirth-related variables. Variable completeness also varies across the data, and while some fields, such as birthweight, have excellent coverage, they may still require data cleaning and validation. For example, implausible values (such as peaks at 0 g and 10,000 g, shown in Figure 4) should be addressed prior to analysis.

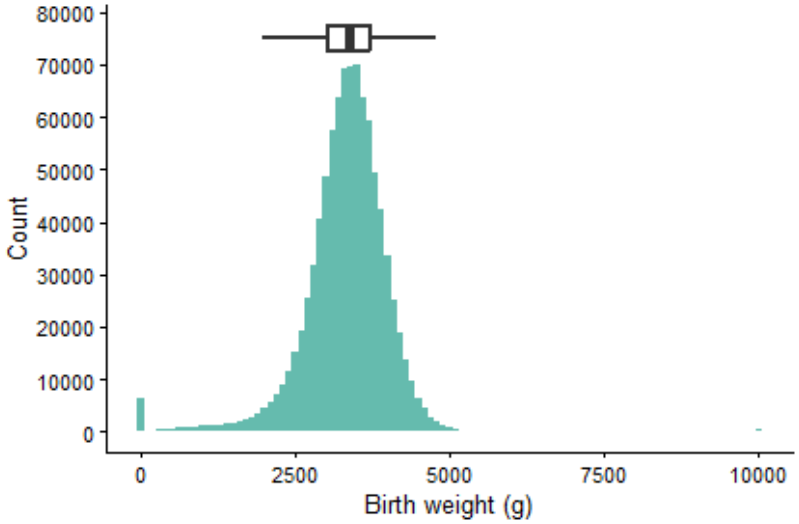


Figure 5 - Histogram and box plot demonstrating the range of birthweights recorded in ADBE.

Data limitations encountered

Changes in coding systems: The classification of stillbirth causes and diagnoses changed from ICD-9 to ICD-10 in 2001. Researchers conducting analyses that span this transition should account for differences in coding structure and definitions, and take care when working across this period.

Multiple births and stillbirth indicators: Indicators for multiple births and stillbirths are recorded as binary flags coded as 1 or NA. This makes it difficult to distinguish between true single births and missing data.

Parental data coverage: Many variables describing parental characteristics, particularly detailed paternal socioeconomic fields, have low completeness rates, which may limit their suitability for some analyses. Researchers are advised to assess variable completeness prior to use.

Lack of maternal linkage: Unlike some other maternity and birth-related data available within SAIL, ADBE does not include a maternal linkage field. This limits maternal analyses and intergenerational research, unless it is linked to other SAIL sources containing the maternal ALF.

References

Office for National Statistics. 2015. Births Statistics: Metadata.

Office for National Statistics. 2025. User Guide to Birth Statistics.

Office for National Statistics. n.d. Impact of the Implementation of IRIS Software for ICD-10 Cause of Death Coding on Mortality Statistics, England and Wales.

Glossary

Abbreviation	Description
ADR UK	Administrative Data Research UK
ALF	Anonymous Linkage Field
DHCW	Digital Health and Care Wales
ESRC	Economic and Social Research Council
GRO	General Register Office
ICD-9	International Statistical Classification of Diseases Version 9
ICD-10	International Statistical Classification of Diseases Version 10
IGRP	Information Governance Review Panel
NHS	National Health Service
NWIS	NHS Wales Informatics Service
ONS	Office for National Statistics
RON	Registration online
RSS	Registration Service Software
SAIL	Secure Anonymised Information Linkage (SAIL) Databank

Disclaimer

This work was produced using administrative data accessed through the SAIL Databank. The use of the data in this work does not imply the endorsement of the SAIL Databank or data owners in relation to the interpretation or analysis.

This work uses research data which may not exactly reproduce National Statistics aggregates.

Acknowledgements

This study uses anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank, which is part of the national population data research infrastructure for Wales. We would like to acknowledge all the data providers who make anonymised data available for research. This study was carried out as part of the ADR Wales programme of work. ADR Wales unites research expertise from Swansea University Medical School and WISERD (Wales Institute of Social and Economic Research and Data) at Cardiff University with analysts from Welsh Government.

This Data Insight has been produced by the ADR Wales Ymateb research team. It provides a snapshot of informative research currently underway at ADR Wales but is not intended to provide a complete picture of work undertaken within this field or the ADR Wales programme of work. The information presented in this Data Insight has been reviewed by ADR Wales colleagues with expertise within this thematic area and accepted to be accurate at the point of publication. Views expressed in this Data Insight are those of the researchers and not necessarily those of ADR Wales partner organisations.

The work is supported by ADR UK (Administrative Data Research UK). ADR UK is a partnership transforming the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve people's lives. ADR UK is an Economic and Social Research Council (ESRC) investment (part of UK Research and Innovation). [Grant number: ES/W012227/1].

Publication contact

Please contact the SAIL Databank via the Helpdesk.