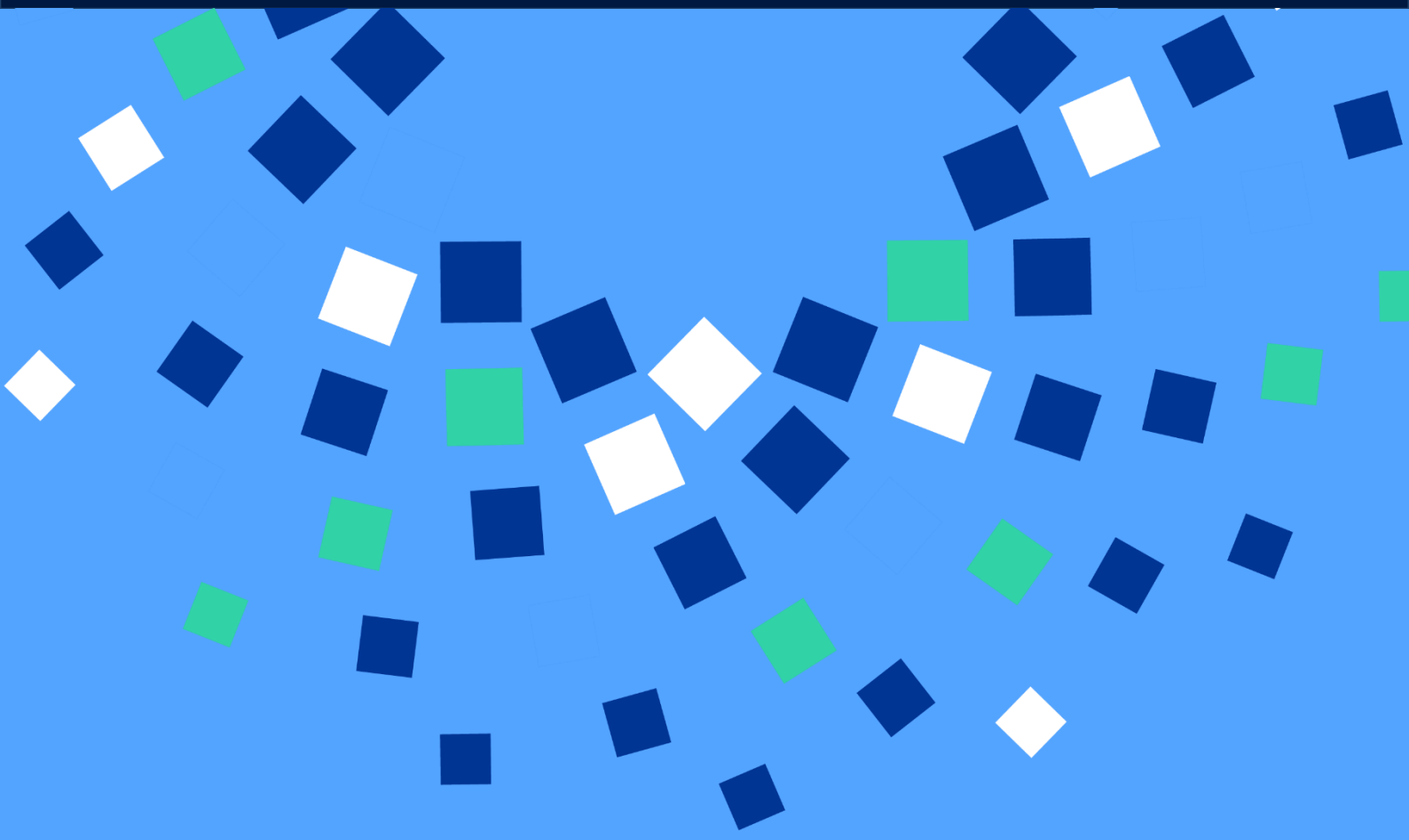
YDG CYMRU | ADR WALES

Data Explained

Patient Episode Database for Wales (PEDW)

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

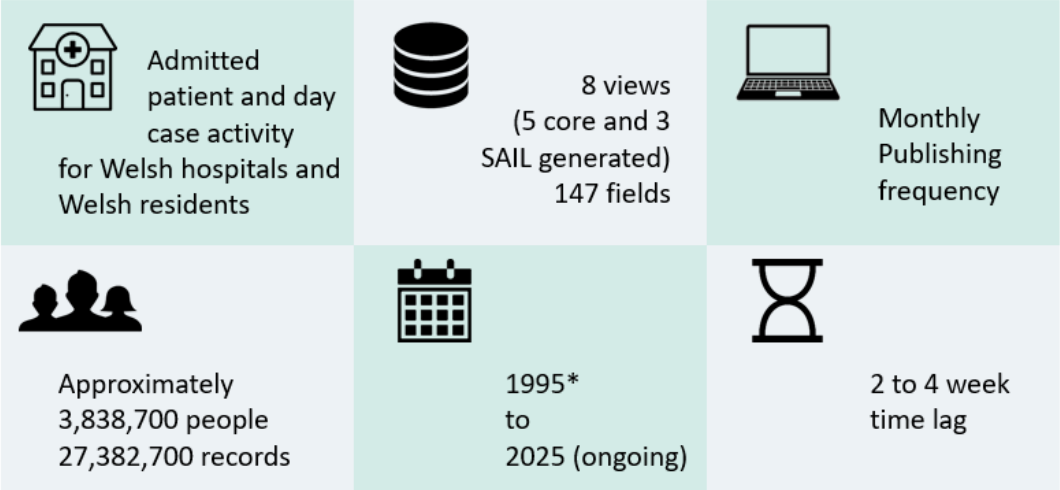


Summary

This Data Explained summarises experiences and learning from working with the Patient Episode Database for Wales (PEDW) 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 Digital Health and Care Wales (DHCW) 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 Patient Episode Database for Wales (PEDW) is an Admitted Patient Care (APC) data source that contains information on secondary care inpatient and day cases across NHS Wales, as well as Welsh residents treated in English Trusts (DHCW, 2020a). It includes data on hospital attendance and clinical information such as diagnoses and procedures (HDRUK, n.d.).

PEDW was implemented in April 1991 (DHCW, 2020a) and was initially developed to record payments for hospital services (HDRUK, n.d.). The data controller for PEDW is Digital Health and Care Wales (DHCW), which is responsible for collecting monthly APC extracts from Welsh NHS sites, as well as downloading data from the NHS Digital Secondary Uses Service (SUS), which provides information on Welsh residents treated in English Trusts (DHCW, 2020a).

How is the data collected?

PEDW data are collected and coded at each hospital for distinct patient episodes. An episode records the time that a patient spends under the care of a single consultant or, following expansion of the data scope in 2024, a nurse. If a patient moves between consultants or specialities, a new episode begins. A hospital spell is defined as the combined duration of admission to the same hospital site. The majority of hospital spells comprise a single episode (approximately 90%) (DHCW, 2024).

Administrative data, such as admission and discharge dates and main speciality of care, are collected from Patient Administrative Systems (PAS). After discharge, clinical coders transcribe handwritten patient notes into standard medical terminology. Diagnoses are recorded using International Statistical Classification of Diseases 10th Revision (ICD-10) codes, and procedures are recorded using Office of Population Censuses and Surveys version 4 (OPCS-4) codes (DHCW, 2020a).

Transcribing handwritten notes takes time and resources, so the administrative record for each finished consultant episode is usually available in PEDW several months before diagnosis and procedure codes are added. While data were historically collected only for finished consultant episodes (FCEs), the scope expanded in 2020 to include unfinished episodes as well.

A timeline of key PEDW events is provided below. This list is not exhaustive but aims to cover the most significant changes relevant for data users.

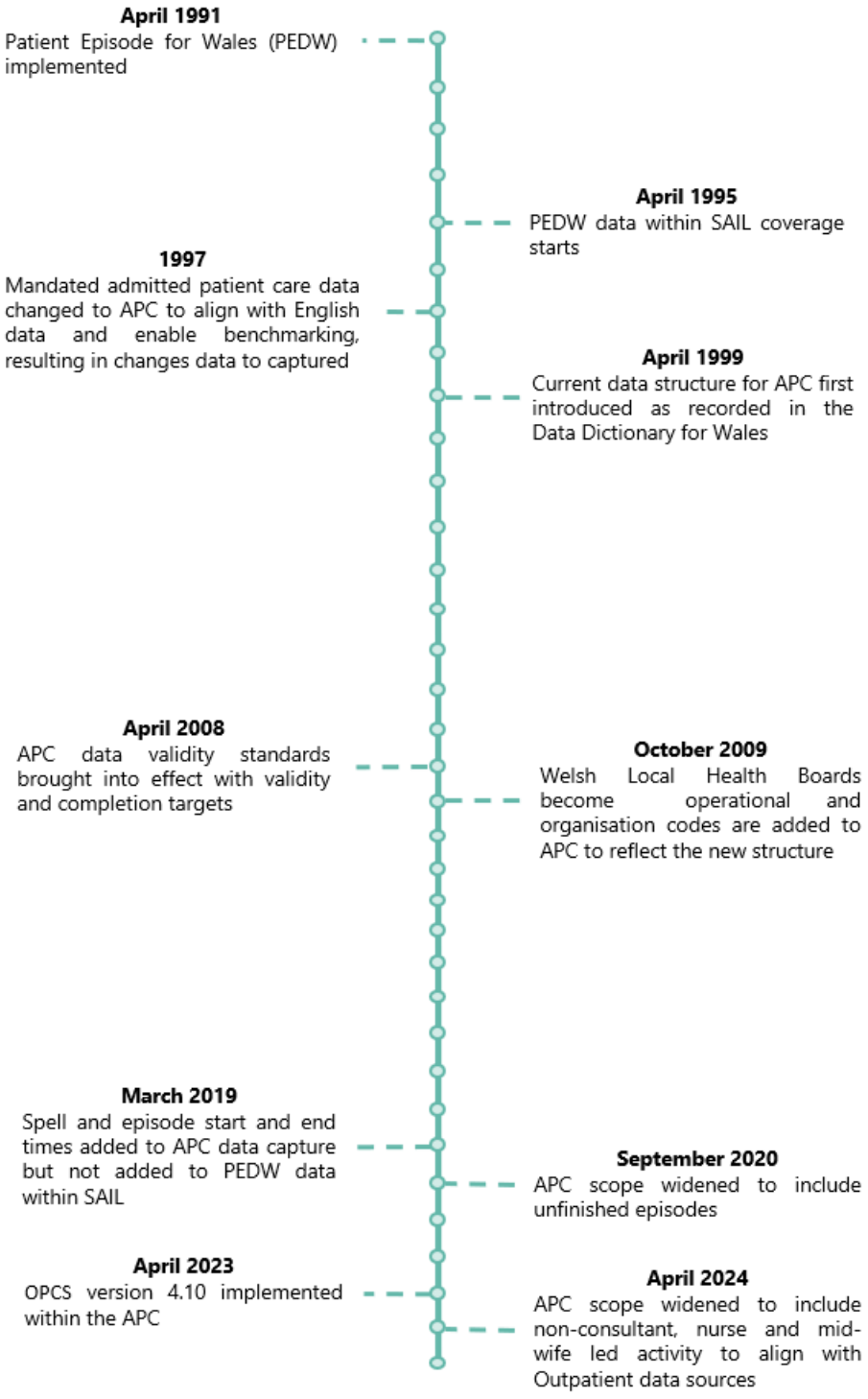


Figure 1 - Timeline of key events in the development of PEDW.

Coverage

PEDW was implemented in 1991, but SAIL data coverage began in April 1995 (HDRUK, n.d.), with relatively few records available before then. However, because the current APC data structure was not introduced until April 1999 (NHS Wales, n.d.), data completeness was likely reduced before then. Information provided by HDRUK indicates an estimated time lag of 2 to 4 weeks for the administrative record to be available.

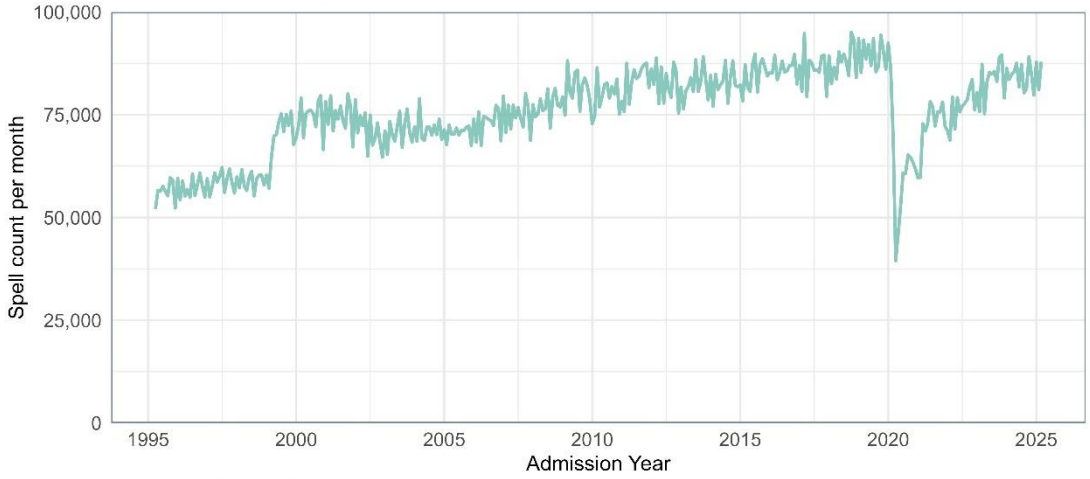


Figure 2 - Count of PEDW spells recorded per month from April 1995 to March 2025.

A further consideration in PEDW coverage is the aforementioned time lag between the availability of administrative records within PEDW and the addition of clinical coding information. When using coding data for research, it is recommended that you explore any impact of data lag and apply a time buffer of 3 to 6 months to allow clinical coding to become available.

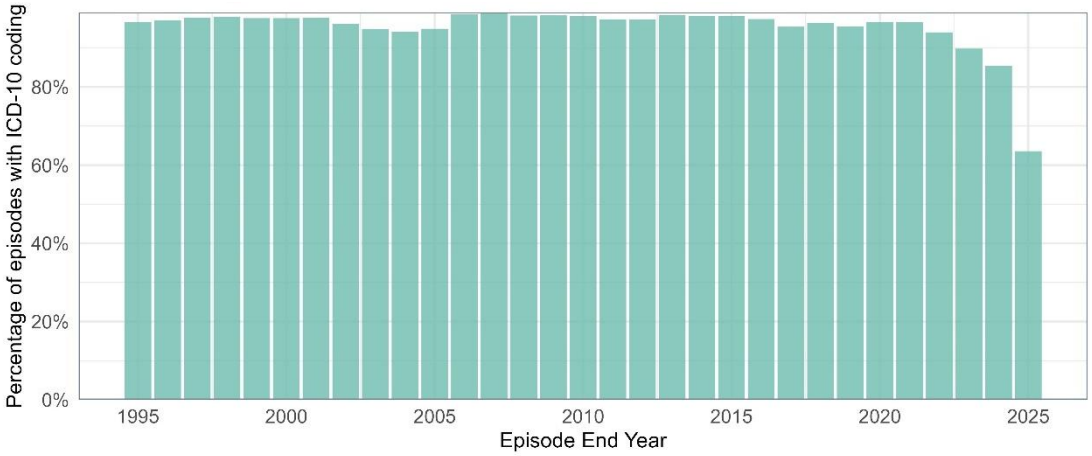


Figure 3 - Percentage of PEDW episodes with ICD-10 coding completion per year, April 1995 to March 2025.

Changes in procedure and recording practices should also be considered when using the data. ICD-10 was mandated for use in Wales from 1995 (DHCW, 2020b), so older records may include ICD-9 codes. Furthermore, ICD-10 is not static; in the UK, it has undergone several iterations, with an updated version issued in 2004, followed by the implementation of version 4 in 2012 and version 5 in 2016 (DHCW, 2020b). Researchers should therefore ensure they link to or refer to time-appropriate reference data.

The provider unit code (prov_unit_cd) identifies the care provider at the time of the admission. This has also been mapped to a current unit provider field (curr_prov_unit_cd), enabling researchers to make like-for-like historical comparisons even when the structure of health care providers has changed.

Format and linkage

The time a patient spends while admitted to the hospital is segmented into distinct periods: superspells, spells and episodes. The superspell (also known as the person spell or continuous person spell) is a derived field that denotes the total duration of a patient's hospital admission from start to finish, which can be across multiple HBs/Trusts in Wales (and England for Welsh residents). A new spell indicates that the patient has moved to a different hospital or care provider within that superspell. A new episode is generated each time a patient changes consultant or main speciality within that spell. Provider spells are defined as continuous periods of inpatient care under a single HB/trust, whereas person spells can take place under many providers.

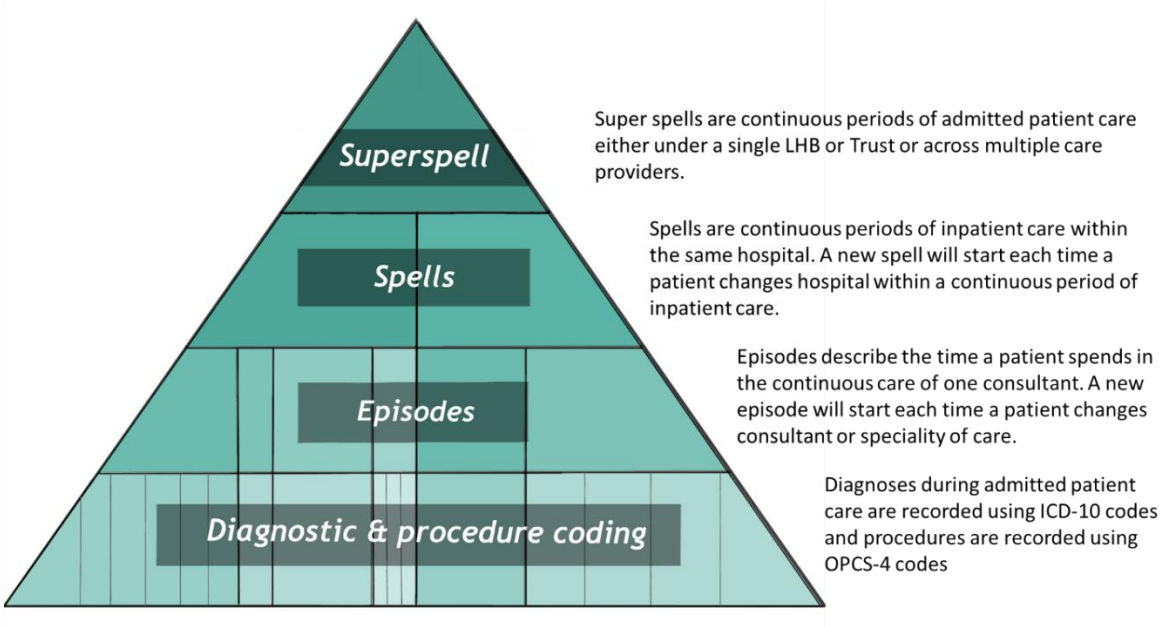


Figure 4 - Visualisation of the structure of inpatient hospital data (adapted from Best et al., 2021).

PEDW is divided into five tables, which reflect this structure: **superspell, spell, episode, diag** and **oper**. Linking these five tables is the most comprehensive way to obtain PEDW data. However, linkage can be resource-intensive, so in some cases it may be more efficient to use the SAIL-derived cleaned tables (see below).

PEDW data tables have no primary key, so unique identifiers for table linkage are generated through a combination of fields. No single common person ID exists across all PEDW data tables, so you may need a series of linkages to connect a person record to the relevant health records. The diagram below indicates the linkage structure and some commonly accessed fields within the PEDW data. ID indicates the fields which together generate a unique ID for that table, with the linkage fields highlighted.

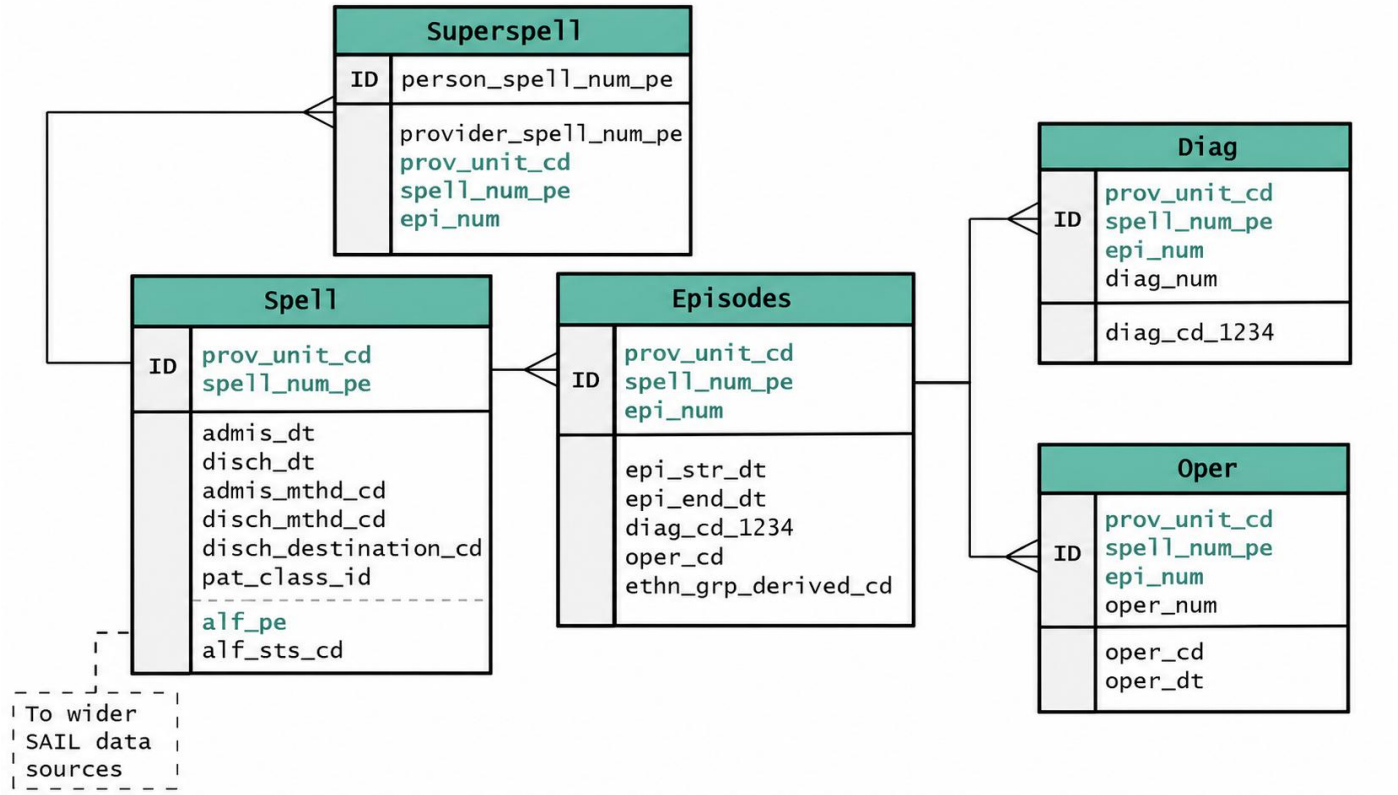


Figure 5 - Visualisation of the data linkage and key fields in the PEDW tables. Note that the green variables are composite keys and they need to be used together.

Spell Table: A unique spell is identified by a combination of the *SPELL_NUM_PE* and *PROV_UNIT_CD* fields. The spell table is the only PEDW table containing the anonymised linkage field (*ALF_PE*), which can be used to link patients to wider SAIL data sources. The spell table contains information about how the patient was admitted, where they were discharged, and the spell start (admission) and end (discharge) dates. Where start and end dates are required for the entire period of admission across multiple care providers, the spells need to be linked to the superspell table.

Superspell Table: The superspell table contains a row per episode, and a unique superspell is identified by the *PERSON_SPELL_NUM_PE* field, under which all spells that are part of the wider superspell are assigned. Superspells are derived from the spell data, and the superspell table is particularly useful where the total time admitted is significant. The superspell table contains sufficient fields to be linked directly to any of the PEDW tables, but it is generally linked to the spell table.

Episode Table: An episode is identified by a combination of the *SPELL_NUM_PE*, *PROV_UNIT_CD* and *EPI_NUM* fields. Episodes are the smallest time unit recorded during a patient's hospital admission and reflect the time spent under a single consultant. Episodes help identify the speciality under which the patient was treated, as well as order diagnoses when multiple episodes exist. The episode table contains primary diagnosis and primary operation codes; however, more detailed information is available in the **diag** and **oper** tables respectively.

Diag Table: Contains details of all the diagnoses made during a hospital admission. A unique diagnosis is identified by a combination of the *SPELL_NUM_PE*, *PROV_UNIT_CD*, *EPI_NUM* and *DIAG_NUM* fields. The **diag** table contains the diagnosis but not the date on which it occurred, so this must be estimated from the episode dates. Diagnoses are recorded in ICD-10 codes.

Oper Table: Contains details of all the operations or procedures conducted during the admission. A unique operation (or procedure) is identified by a combination of the *SPELL_NUM_PE*, *PROV_UNIT_CD*, *EPI_NUM* and *OPER_NUM* fields. The **oper** table contains the operation code in OPCS-4 format and the date the procedure started.

SAIL derived cleaned tables: The derived cleaned tables are generated by the cleaning routine (HDRUK, n.d.). The derived, cleaned tables all contain the *ALF_PE* field, allowing them to be directly linked to the wider SAIL data containing *ALF_PE* if desired, potentially reducing the number of joins required and thereby improving performance. The derived, cleaned tables also contain sufficient data-linkage fields to be directly linked to each other.

There are three derived cleaned tables: the **pedw_admissions** table provides a single row per superspell (or person spell); the **pedw_single_diag** table provides a single row per diagnosis; and the **pedw_single_oper** table provides a single row per operation or procedure. However, the derived cleaned tables sacrifice some detail. Admission start dates and end dates are available only for the superspell admission and discharge. Therefore, where date nuance is required, the original PEDW tables may be more useful. Similarly, if you need fields not included in the derived cleaned tables, you may still need to link back to the original PEDW tables, negating any previously achieved efficiencies.

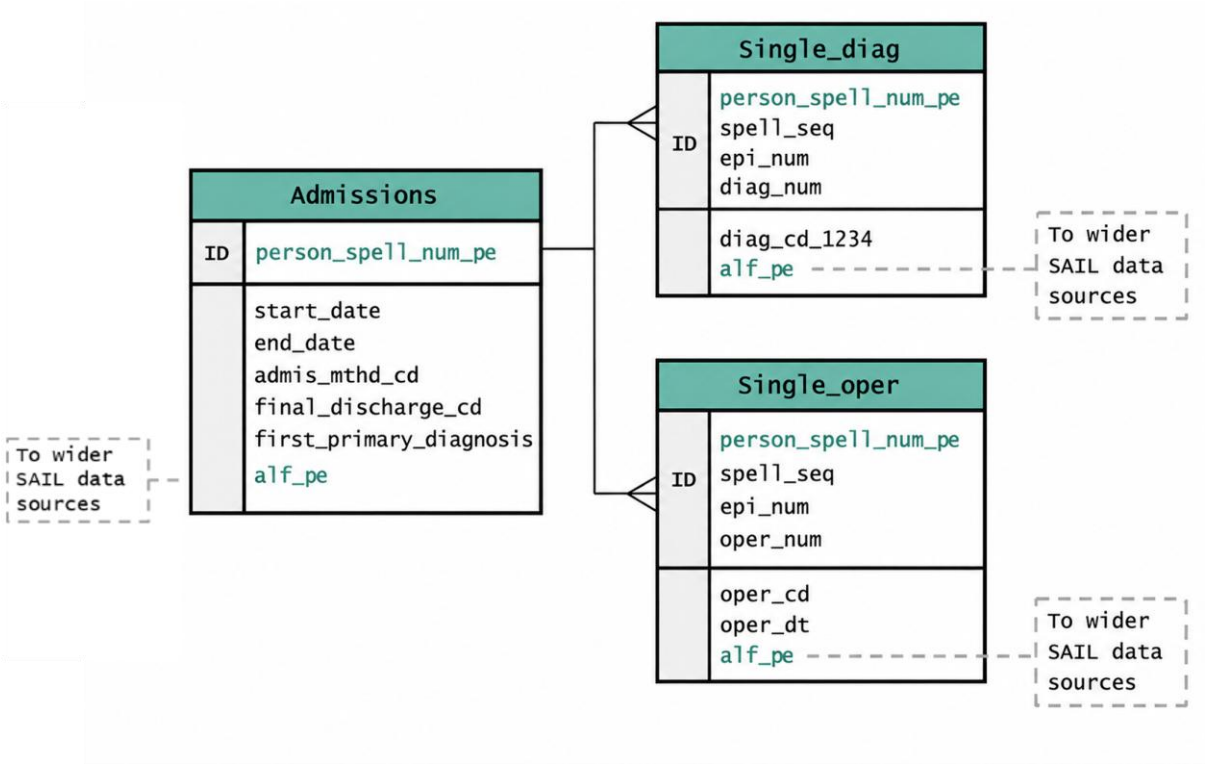


Figure 6 - Visualisation of the data linkage and key fields in the PEDW SAIL derived cleaned tables.

What can the data be used for?

PEDW is an extensive source of secondary care data in Wales and has been widely used in research and epidemiology. It includes elective admissions, i.e. planned hospitalisation where care or treatment is scheduled in advance, emergency admissions, i.e unplanned hospitalisations resulting from acute medical emergencies, accidents, or urgent referrals requiring immediate care, maternity admissions, i.e. admissions of pregnant or recently pregnant women to maternity wards, and day case admissions, i.e planned, recurring elective admissions part of an ongoing therapeutic regimen.

Further insight can be gained by linking to wider data sources held within the SAIL Databank, such as primary care, social care, and demographic data. Examples of what PEDW can be used for include:

- Identifying hospital diagnoses, operations and length of stay for Welsh residents in Wales and England.
- Establishing the incidence and prevalence of disease.
- Identifying the presence of conditions and comorbidities of interest within a cohort.
- Mapping a patient's journey through the healthcare system.

ICD-10 coding is internationally utilised, enabling morbidity data to be compared across countries and over time and facilitating the statistical study of disease. The World Health Organisation (2019) publishes details of the ICD-10 chapters and the codes for each diagnosis.

OPCS-4 codes are the clinical coding standard for recording surgical procedure and intervention information across the NHS, although there is some local variation between England, Wales, Scotland and Northern Ireland (NHS England, 2024). OPCS-4 codes can be used for research and epidemiology; details are available on the Classifications Browser published by the NHS (NHS, n.d.).

Much of the data recorded in PEDW fields is coded; therefore, reference data is required to establish its meaning. SAIL provides many reference data lookup tables in the UKHD schema, and coding generally reflects wider standards available in the NHS Wales Data Dictionary (NHS Wales, n.d.). PEDW metadata is available on the HDR UK Innovation Gateway (n.d.).

Data limitations encountered

DHCW produce quarterly and annual statistical reports summarising multiple aspects of the PEDW data ("Quarterly Data Tables - Statistical Reports," 2023).

Diagnosis date: PEDW does not contain the diagnosis date. The start and end dates of the episode in which the diagnosis was recorded are available; however, it is not possible to determine precisely when within this episode the diagnosis occurred. This presents an issue for any research requiring the specific ordering of events, or for research that requires distinction between hospital- and community-acquired conditions.

Admission time: PEDW does not include admission time, limiting the ability to distinguish hospital-acquired from community-acquired conditions, many of which are defined in hours from admission, such as Hospital Acquired Pneumonia (HAP), which must occur at least 48 hours after admission (Metlay et al., 2019). The APC data source has recorded spell and episode start times since 2019 (Jones, n.d.), but this information has not been added to the PEDW SAIL data due to current privacy protection standards.

Treatment in England: PEDW contains data for Welsh residents treated in Wales and England, however, according to DHCW, "It is possible that not all the data on Welsh residents treated in England is received through NHS SUS extracts" (DHCW, 2024).

Coding lag: Delay in completing diagnostic and procedural coding is a known issue for PEDW data, and DHCW produces reports on data quality and validity, which can be accessed via the NHS Wales Intranet (DHCW, 2024). Lag in clinical coding for PEDW is also covered in the data insight (Best et al., 2021). The majority of clinical coding would usually be expected to be available within approximately 3 months of the episode end; however, at the time of writing, staffing issues within the NHS Clinical Coding Service were causing additional delays. DHCW cautions about this, highlighting that the number of episodes with assigned classification codes is significantly lower in the 2025/26 financial year than in previous years, and that this is likely to affect any all-Wales analysis (DHCW, 2024). While coding for the 2023/24 and 2024/25 financial years was similarly incomplete at the time of the previous version of this report, these years are now more up to date.

ALF_PE completeness: Within SAIL, the *ALF_PE* enables linkage between data sources, so records without an *ALF_PE* are commonly excluded for SAIL epidemiological research purposes. In earlier years of coverage, the proportion of spells with an *ALF_PE* is low, which reduces the usefulness of records prior to 1998. Because the PEDW structure also changed in 1999, researchers should exercise additional caution when using data before this date.

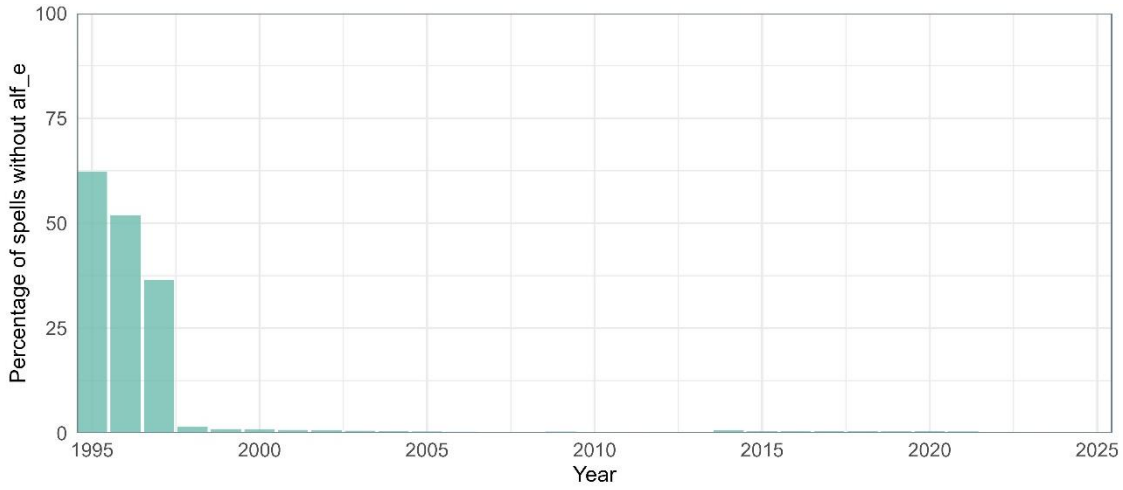


Figure 7 - Percentage of spells per year with no anonymous linkage field (ALF_PE). ***From 1999, the percentage of spells without an ALF_PE is <2%.***

References

Best, V., Akbari, A., Bedston, S., Lowthian, E., Lyons, J., & Torabi, F. (2021). Informing Better use of Welsh Hospital Admissions Data in the SAIL Databank. *ADRUK*. https://www.adruk.org/fileadmin/uploads/adruk/Documents/Data_Insights/Data_Insight_Patient_Episode_Database_for_Wales_Clinical_Coding_January_2022.pdf

DHCW. (2020a, November 12). *Archived PEDW Data Online*. Digital Health and Care Wales. https://dhcw.nhs.wales/data/statistical-publications-data-products-and-open-data/archived-pedw-data-online/

DHCW. (2020b, December 7). *ICD-10*. Digital Health and Care Wales. https://dhcw.nhs.wales/data/information-standards/clinical-classifications-and-terminology-standards/about-clinical-classifications-and-coding/icd-10/

DHCW. (2024, January 10). *Annual Admitted Patient Care (formerly PEDW) Data Tables—Notes & Definitions 2024 /25*. https://dhcw.nhs.wales/data/statistical-publications-data-products-and-open-data/hospital-data/hospital-admissions-annual-tables/notes-and-definitions/

HDRUK. (n.d.). *Dataset—Health Data Research Gateway*. Retrieved 4 February 2025, from https://healthdatagateway.org/en/dataset/318

Jones, P. (n.d.). *Ministerial / Official Letter: N/A Subject: APC data set – Time fields*.

Metlay, J. P., Waterer, G. W., Long, A. C., Anzueto, A., Brozek, J., Crothers, K., Cooley, L. A., Dean, N. C., Fine, M. J., Flanders, S. A., Griffin, M. R., Metersky, M. L., Musher, D. M., Restrepo, M. I., & Whitney, C. G. (2019). Diagnosis and Treatment of Adults with Community-acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America. *American Journal of Respiratory and Critical Care Medicine*, *200*(7), e45–e67. https://doi.org/10.1164/rccm.201908-1581ST

NHS. (n.d.). *Classifications Browser OPCS-4.10*. Retrieved 31 October 2025, from https://classbrowser.nhs.uk/#/book/OPCS-4.10

NHS England. (2024, April). *National Clinical Coding Standards OPCS-4*. https://classbrowser.nhs.uk/ref_books/OPCS-4.10_NCCS-2024.pdf

NHS Wales. (n.d.). *Data Set Structure*. NHS Wales Data Dictionary. Retrieved 31 October 2025, from https://www.datadictionary.wales.nhs.uk/#!WordDocuments/datasetstructure21.htm

Glossary

Abbreviation	Description
ADR UK	Administrative Data Research UK
ALF	Anonymised Linkage Field
APC	Admitted Patient Care
DHCW	Digital Health and Care Wales
FCE	Finished Consultant Episode
ICD-9/10	International Statistical Classification of Diseases Version 9/10
IGRP	Information Governance Review Panel
HAP	Hospital Acquired Pneumonia
HB	Health Board
HDRUK	Health Data Research UK
OPCS-4	OPCS Classification of Interventions and Procedures 4th Edition
PAS	Patient Administrative System
SAIL	Secure Anonymised Information Linkage (SAIL) Databank
SUS	Secondary Uses Service
WHO	World Health Organisation

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.