Administrative Data
Research Centre
Wales

An ESRC Data
Investment

Better Knowledge
Better Society

Data Brief 3: An Overview of
Administrative Health Data
within the SAIL Databank

September 2015
Available in Welsh on request

The Secure Anonymised Information Linkage (SAIL) Databank[1] contains a number of administrative health datasets derived from a variety of sources, such as General Practice records, Annual District Birth and Death Extracts, and various types of hospital data. These datasets contain key information relating to the health and well-being of the population in Wales. They can be linked in order to address important research questions that could lead to both a better understanding of health in Wales and how health impacts upon other socio-economic outcomes.

This data brief provides an overview of the administrative health datasets within the SAIL Databank.

Mae'r System Ddiogel ar gyfer Cysylltu Gwybodaeth Ddienw yn cynnwys nifer o Setiau Data Gweinyddol sy'n deillio o nifer o ffynonellau, megis cofnodion Meddygon Teulu, Crynodebau Blynyddol o'r Cofrestri Blynyddol Rhanbarthol o Enedigaethau a Marwolaethau, a gwahanol fathau o ddata o Ysbytai. Mae'r setiau data hyn yn cynnwys gwybodaeth allweddol yn ymwneud ag iechyd a lles poblogaeth Cymru. Mae modd eu cysylltu er mwyn mynd i'r afael â chwestiynau ymchwil pwysig a allai arwain at ddealltwriaeth well o iechyd yng Nghymru ac at ddealltwriaeth well o effaith iechyd ar ganlyniadau economaidd-gymdeithasol eraill.

Mae'r crynodeb data hwn yn rhoi trosolwg o'r Setiau Data Gweinyddol o fewn y System.

Introduction

Patients come in to contact with the National Health Service (NHS) in Wales around 20 million times each year; with roughly 700,000 outpatient appointments, more than 600,000 inpatient and day-case appointments, and more than 1 million people attending Accident and Emergency (A&E) departments[2]. Whenever a person attends their GP surgery or is admitted to hospital, a record of that event is generated. This large amount of electronic health data that is routinely collected and stored about patients could benefit health-related research, service planning, policy evaluation, and ultimately improve quality of care.

The development of the Secure Anonymised Information Linkage (SAIL) Databank[1] has resulted in this abundance of administrative health data being maximally utilised, without compromising patient identity. SAIL is an initiative developed by Swansea University Medical School and receives core-funding from Health and Care Research Wales of the Welsh Government. SAIL receives a number of health and social care datasets relating to the population in Wales, from various organisations. SAIL has expanded the databank, both in types of dataset held and in geographical coverage.

Primary Care GP Dataset

There are currently a total of 462 GP practices actively treating patients in Wales. Of those practices 350 (76%) have agreed to include their data within the GP dataset held in SAIL. The available data covers the period from January 2000 to the present, but this varies by practice. Individual patient level clinical information is securely extracted (using Audit+) from the practices' clinical systems. This system captures the signs, symptoms, test results, diagnoses, prescribed treatment, referrals for specialist treatment and social aspects relating to the patients' home environment. The majority of the data is entered by the clinician during the patient consultation. However, other records are also created, e.g. interactions with others in the practice team, repeat prescriptions, and some test results from secondary care systems.

There are no standard rules for recording data within primary care clinical information systems. Therefore, each individual clinician can record information in their own way. GP data is coded in a hierarchical system called Read Coding. To improve the consistency of coding, standards have been agreed for certain conditions monitored by the Quality and Outcomes Framework (QOF). Events tend to be recorded when relevant to a specific primary care interaction rather than routinely. In SAIL it is not possible to distinguish what kind of interaction generated a particular record. For example, a face-to-face visit to a GP, a test carried out by a practice nurse, or printing of a repeat prescription. Each single piece of information recorded in primary care creates a record, so two prescriptions and a blood pressure reading would create three records for a patient on the same day.

The GP dataset has previously been linked to other datasets to examine a number of important research questions. These have ranged from looking at differences in blood pressure level readings in Primary and Secondary care[3], to the rate of acute myocardial infarction or stroke among patients with ankylosing spondylitis[4]. The GP dataset also allows for examination within individual health boards and at GP cluster level. For example, research has used GP data from the Abertawe Bro Morgannwg University Health Board to examine chronic pulmonary disease in relation to flu vaccine uptake and emergency admissions resulting from respiratory disease during the flu period[5].

Table 1: Number and population coverage (count & percentage) of GP Surgeries in Wales currently signed up to SAIL (as at 30th June 2015).

Health Board	Number of GP practices	Number of practices signed up to SAIL (%)	Total population coverage (%)
ABMU	76	76 (100%)	542,420 (100%)
Aneurin Bevan	88	55 (63%)	396,781 (66%)
Betsi Cadwaldr	114	82 (72%)	531,428 (75%)
Cardiff & Vale	66	53 (80%)	420,971 (85%)
Cwm Taf	47	34 (72%)	235, 173 (79%)
Hywel Dda	54	43 (80%)	319,386 (77%)
Powys	17	7 (41%)	55,739 (41%)
Totals	462	350 (70%)	2,501,898 (83%)

[1] www.saildatabank.com

[2] Welsh Government Statistics: http://gov.wales/topics/health/nhswales/about/?lang=en

[3] Brooks, C.J., Tang, T.S., Ford, D.V., Lyons, R.A., Price, D.E., Bain, S.C., & Stephens, J.W. (2009). Are blood pressure levels taken during a secondary care diabetic clinic likely to be higher than when measured in primary care? *Primary Care Diabetes, 3(3)*, 193-195. http://www.sciencedirect.com/science/article/pii/S1751991809000461

[4] Brophy, S., Cooksey, R., Atkinson, M., Zhou, S.M., Husain, M.J., Macey, S., Rahman, M.A., & Siebert, S. (2012). No increased rate of acute myocardial infarction or stroke among patients with ankylosing spondylitis – A retrospective cohort study using routine data. *Seminars in Arthritis and Rheumatism, 42(2)*, 140-145. http://www.sciencedirect.com/science/article/pii/S0049017212000327

[5] Gibbon, R., May, L., Humphreys, C., Price, T., Hughes, R., Holloway, M., Francis, I., Lester, N., & Puscas,I. (2012). *Chronic obstructive pulmonary disease in the Abertawe Bro Morgannwg area report IIb:COPD severity, flu vaccination and emergency admissions.* Public Health Wales Observatory. http://www.saildatabank.com/media/19317/20120725_copdinabmu_fluvaccpartiib_lmrgtp_v1afinal.pdf

ONS Births and Deaths Extracts

The NHS Wales Informatics Service (NWIS) provides the Annual District Birth Extract (ADBE) and the Annual District Death Extract (ADDE) on behalf of the Office for National Statistics (ONS). The datasets contain registry data of all registered births and deaths in Wales from 2003 to the present. Records from the ADBE show there are approximately 35,000 births per year in Wales with data recorded on an individual person level base. The key variables within the ADBE include: date and place of birth, gender of the baby and mother's and father's place of birth.

The Annual District Death Extract records individual person level information. There are approximately 32,000 deaths per year in Wales. The key variables within the ADDE include: Cause of death, age at death, gender, and area of residence. The ADDE has been used in a variety of research studies covering a range of areas relating to death rates in Wales. For example, one study examining the incidence of hospitalisation and case fatality for upper gastrointestinal bleeding demonstrated an increase in mortality among people who were admitted on weekends and public holidays compared to those admitted during the week[6].

Patient Episode Database for Wales (PEDW)

This is the national repository for all episodes of inpatient and day-case activity in all NHS Wales hospitals. All hospitals are required to collect and code standardised data. This data is downloaded on a monthly basis to the Patient Administration Systems (PAS) database. These are then collected by NWIS, where they are validated and merged in to the main PEDW database. NWIS also receives details of Welsh patients treated in England through a mechanism known as the NHS Switching Service. This switching service also provides details of English patients treated in Wales to the English equivalent system called Hospital Episode Statistics (HES). The information from the NHS Switching Service is also fed in to the PEDW.

The PEDW contains attendance and clinical information for all hospital admissions, which includes all finished consultant episodes of inpatient and day-case care. A finished consultant episode is defined as a completed 'unit' of care under one consultant. Each episode has a provision for a number of diagnoses and operative procedure codes to be recorded. The PEDW utilises the International Classification of Diseases (ICD) diagnostic codes, and the OPCS-4 classification of surgical operations.

There are approximately 950,000 hospital admissions and over 1 million episodes of hospital care per year. The data held in the PEDW can provide information regarding both health service utilisation and the incidence and prevalence of disease. However, when there is no definitive diagnosis within a hospital episode the main symptom, abnormal finding or problem is recorded as the primary diagnosis. Furthermore, caution must be taken when utilising the PEDW data as counts will vary depending on the number of diagnosis fields used (primary only versus using all fields), and there may also be regional variation in coding. There are also a number of different ways the PEDW can be explored, such as individual episodes of care, admissions, discharges, periods of continuous care (group of episodes) and patients or procedures. The validation process led by the Corporate Health Improvement Programme and implemented by NWIS is aiming to address some of these inconsistencies.

The PEDW has been used in a variety of research studies covering a range of areas relating to hospital episodes in Wales. One study utilised incidences of hospitalisation from PEDW in order to examine the influence of deprivation and air pollution on serious asthma[7], while another study has examined whether PEDW data can be used to support a national audit of the management of inpatients with inflammatory bowel disease[8].

Comorbidity Index

Comorbidity refers to the presence of multiple diseases or disorders in a patient, which may be interrelated or independent of each other. An algorithm has been developed by the SAIL team that can be run against the PEDW in order to produce a comorbidity index for each patient based on the condition a researcher wishes to look at. This algorithm is based on the Charlson Comorbidity Index (CCI)[9], which aims to predict the mortality of a patient who may have a range of co-morbid conditions such as heart disease or cancer. The index contains 17 categories of comorbidity with each condition being assigned a score between 1 and 18 depending on the risk of dying associated with each condition. Scores are summed to provide a total score to predict mortality. Higher scores indicate greater comorbidity.

Clinical conditions and associated scores are as follows:
Diabetes complications: 1 point
Paraplegia: 1 point
HIV/AIDS: 2 points
Diabetes: 3 points
Connective tissue disorder, pulmonary disease: 4 points
Acute Myocardial infarction: 5 points
Peripheral vascular disease: 6 points
Mild or moderate liver disease, cancer: 8 points
Peptic ulcer: 9 points
Renal disease: 10 points
Cerebral vascular accident: 11 points
Congestive heart failure: 13 points
Dementia, metastatic cancer: 14 points
Severe liver disease: 18 points

There have been many versions of the CCI and many other indices to measure comorbidity. The SAIL team have used the version described by Bottle and Aylin[10] but using the weights listed by the Health and Social Care Information Centre (HSCIC) Clinical Indicators Team[11]. There is no access to any data on HIV/AIDS, thus there is no data in this section. The weights in Charlson's original version are very different from the ones now widely used due to improvements in treatment. The score is helpful for a physician to decide how aggressively to treat a condition. For example, a patient may have cancer, heart disease and diabetes. The comorbid conditions may be so severe that the cost and risk of cancer treatment would outweigh any short-term benefits.

Emergency Department Data Set (EDDS)

NWIS provides this dataset which contains administrative and clinical information for all activity/attendances at all A&E departments and Minor Injury Units (MIU) in Wales. Whilst all attendances irrespective of residence are recorded, Welsh residents attending emergency departments in hospitals in England are not included. The EDDS also includes the All Wales Injury Surveillance Systems (AWISS) dataset[12] which collects data on all injuries from A&E departments. The data are collected and coded at each hospital where they are electronically transferred to NWIS to be validated and merged in to the main database. Diagnoses, investigations, and treatments are coded using broad categories.

The EDDS contains information on attendances from 2009 to the present and includes records on approximately 750,000 attendances per year. The data held in the EDDS can provide information regarding both health service utilisation and also the incidence and prevalence of certain events. When counting activity it is possible to count individual attendances (new and/or follow-up), individual attendances leading to an admission and persons attending (irrespective of how many times they have attended within a given period). Whilst validity checks are undertaken by the Health Boards and NWIS, researchers need to be mindful of local variations in coding practice. Despite these challenges, the EDDS has been used in previous research studies covering a range of areas relating to emergency data in Wales. For example, one study has used the EDDS to examine whether a change in the amount of alcohol outlets within local authorities had any impact on alcohol consumption levels and alcohol-related harm to health[13].

[6] Button, L.A., Roberts, S.E., Evans, P.A., Goldacre, M.M., Akbari, A., Dsilva, R., Macey, S., & Williams, J.G. (2011). Hospitalized incidence and case fatality for upper gastrointestinal bleeding from 1999 to 2007: a record linkage study. *Alimentary Pharmacology & Therapeutics, 33 (1)*, 64-76. http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2036.2010.04495.x/abstract

[7] Roberts, S.E., Button, L.A., Hopkin, J.M., Goldacre, M.J., Lyons, R.A., Rodgers, S.E., Akbari, A., & Lewis, K.E. (2012). Influence of social deprivation and air pollutants on serious asthma. *European Respiratory Journal, 40(3)*, 785-788. http://erj.ersjournals.com/content/40/3/785.full

[8] Roberts, S.E., Williams, J.G., Cohen, D.R., Akbari, A., Groves, S., & Button, L.A. (2011). Feasibility of using routinely collected inpatient data to monitor quality and inform choice: a case study using the UK inflammatory Bowel Disease Audit. *Frontline Gastroenterology, 2 (3)*, 153-159. http://fg.bmj.com/content/early/2011/04/05/fg.2009.000208.short

[9] Charlson, M.E., Pompei, P., Ales, K. L., MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases, 40 (5)*, 373-383. http://www.sciencedirect.com/science/article/pii/0021968187901718

[10] Bottle, A., & Aylin, P. (2011). Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices. *Journal of Clinical Epidemiology, 64*, 1426-1433. http://www.sciencedirect.com/science/article/pii/S0895435611001211

[11] Health & Social Care Information Centre (HSCIC) Clinical Indicators Team (2011). Methodology: Charlson comorbidity index. 2nd August 2011, Version 1.1.

[12] Lyons, R.A., Jones, S., Kemp, A., Sibert, J., Shepherd, J., Richmond, P., Bartlett, C., & Palmer, S.R. (2002). Development and use of a population based injury surveillance system: the All Wales Injury Surveillance System (AWISS). *Injury Prevention, 8 (1)*, 83-86. http://injuryprevention.bmj.com/content/8/1/83.short

[13] Fone, D., Dunstan, F., White, J., Webster, C., Rodgers, S., Lee, S., Shiode, N., Orford, S., Weightman, A., Brennan, I., Sivarajasingam, V., Morgan, J., Fry, R., & Lyons, R. (2012). Change in alcohol outlet density and alcohol-related harm to population health (CHALICE). *BMC Public Health, 12: 428*. http://www.biomedcentral.com/1471-2458/12/428/

Outpatients Dataset (OPD)

NWIS provide this dataset which contains attendance information for all NHS Wales hospital outpatients' appointments. The data are collected and coded at each hospital. Administrative information such as speciality of care, appointment date and attendance status is collected from the central Patient Administration System (PAS). There are approximately 4,500,000 appointments per year. The dataset contains individual person level data and is available from 2004 to the present. The dataset also contains all scheduled outpatient appointments including those where the patient failed to attend.

Previous research studies relating to hospital outpatient activity in Wales have already been conducted using the OPD. For example, one study has linked outpatient data with a number of other health datasets, questionnaire data, and clinical studies in order to produce a single data resource for use in health economic analyses and longitudinal tracking of patient journeys through the healthcare system[14].

National Community Child Health Database (NCCHD)

The child health system in Wales includes the registration of births and the monitoring of child health examinations and immunisations. The NCCHD brings together data from local Child Health System databases held by NHS Health Boards which are used to administer child immunisation and health surveillance programmes. There are approximately 35,000 births per year and some 500,000 vaccinations each year in Wales. The

dataset contains individual level data on all children born, resident, or treated in Wales and is available from 1987 to the present. It contains information on where and when babies were born, their birth weight and whether they were pre-term. It also includes information about their mothers such as their age and whether they breastfed the baby.

The NCCHD is used to provide key statistics on the incidence of births in Wales[15][16] in order to inform the delivery of maternity services and strategy in Wales. Research studies have already used the NCCHD to address important research questions relating to child health. For example, one study has linked the NCCHD with data from the Welsh Cervical Screening Programme (CSW, see below) to examine whether there are adverse birth outcomes following treatment for precancerous changes to the cervix[17].

Cancer Datasets

Welsh Cancer Intelligence and Surveillance Unit (WCISU)

This is the National Cancer Registry for Wales. Its primary role is to record, store and report on all incidence of cancer for the resident population of Wales wherever they are treated. The electronic database, which holds individual person level records from 1972, contains in the region of 686,000 records. The WCISU collects data about occurrences of cancer in Welsh residents via direct or indirect submissions from Welsh hospitals. It can provide high quality information about cancer incidence, survival and mortality and can monitor trends in cancer. Staging of malignant melanoma, breast, colorectal and

cervical cancers started in 2001. Staging for all other cancers started in 2010. Treatment information started in 1995. The database can be used to facilitate the planning of cancer services for prevention, diagnosis, cure and care. Researchers should be aware that revisions to the ICD classification have resulted in changes in the coding of cancer during the period covered by the dataset.

Public Health Wales Cancer Screening Datasets

Public Health Wales collates a variety of administrative and clinical information related to different cancer screening schemes in operation in Wales:

▶ Bowel screening test kits are currently offered to men and women aged between 60 and 74 in Wales. The Bowel Screening Wales dataset contains information on approximately 280,000 invitations for bowel cancer screening and approximately 140,000 screening test records per year. The data is held at an individual person based level and is available from the end of October 2008 to the present.

▶ The Breast Test Wales dataset contains administrative and clinical information for routine breast screening; currently offered to women aged between 50 and 70 who are resident in Wales, with older women being eligible for self-referral. The dataset contains approximately 5,500 assessments and 110,000 screening test records per year. The data is held at an individual person based level and is available from February 1989 to the present.

▶ The Cervical Screening Wales (CSW) dataset contains administrative and clinical information for cervical screening; currently offered to women aged between 24 and 64 who are resident in Wales. Up until 2013 the initial invitation was sent to women aged 20. The data contains approximately 335,000 invitations, 225,000 screening tests and 25,000 assessment records per year. The data is held at an individual person based level and is available from January 1990 to the present for invitation and screening test records and from April 2011 to the present for assessment data.

Previous research has made use of the CSW data in order to investigate the uptake of cervical screening following the implementation of the HPV vaccination programme for a catch-up group of women who were thought to still benefit from the vaccine despite being older than the required age[18].

The Congenital Anomaly Register and Information Service (CARIS)

The CARIS collects information about any fetus or baby who has, or is suspected of having, a congenital anomaly and whose mother is normally a resident in Wales at the

time of birth. The CARIS defines an anomaly as involving structural, metabolic, endocrine, or genetic defects, present in the child/fetus at the end of pregnancy, even if it is not detected until later. It includes babies in whom anomalies are diagnosed at any time from conception to the end of the first year of life. Data collection commenced on the 1st of January 1998 and includes any baby where pregnancy ended after this date[19].

The CARIS aims to collect reliable data about congenital anomalies that can then be used to help build up and monitor the picture of congenital anomalies in Wales, assess interventions intended to help prevent or detect congenital anomalies, plan and co-ordinate provision of health services for affected babies and children, and assess possible clusters of birth defects and their causes.

The CARIS uses a multi-source data collection method using a wide range of sources within the NHS. This ranges from antenatal ultrasound, clinical letters, post-mortems and laboratory results. The CARIS also accesses a number of databases including the SHIRE (a Medical Genetics Database), the PEDW, the NCCHD, and the Paediatric Cardiology Database. Medical records are accessed to confirm, validate and add further details to the information already collected.

The CARIS dataset contains individual level information on approximately 1,500 babies per year. Data linkage based on the details of the mother can be performed. Not all babies included in the CARIS dataset will be born with a congenital anomaly and other checks need to be performed to distinguish those children born with congenital anomalies from those born with other significant conditions. The CARIS contains information on the Anatomical Therapeutic Chemical Classification (ATC codes), which is used to code medications taken by the mother during pregnancy, and allows for investigation in to the possible associations between particular types of drugs and any anomalies. The CARIS release the dataset annually and the SAIL anonymised version is updated accordingly. The CARIS has been previously used as part of a research study to examine the risk of major congenital anomalies following in utero medication exposure[20].

[14] Husain, M.J., Brophy, S., Macey, S., Pinder, L.M., Atkinson, M.D., Cooksey, R., Phillips, C.J., Siebert, S. (2012). HERALD (Health Economics using Routine Anonymised Linked Data). *BMC Medical Informatics and Decision Making, 12:24.* http://www.biomedcentral.com/1472-6947/12/24/

[15] National Community Child Health Database assists Welsh Government Statistics. Health in Wales News archive. http://www.wales.nhs.uk/news/28291

[16] Births: Data from the National Community Child Health Database. Welsh Government Statistics. http://gov.wales/statistics-and-research/births-national-community-child-health-database/?lang=en

[17] Reilly, R., Paranjothy, S., Beer, H., Brooks, C., Fielder, H.M., Lyons, R.A. (2012). Birth outcomes following treatment for precancerous changes to the cervix: a population-based record linkage study. *BJOG, 119,* 236-244. http://onlinelibrary.wiley.com/doi/10.1111/j.1471-0528.2011.03052.x/full

[18] Beer, H., Hibbitts, S., Brophy, S., Rahman, M.A., Waller, J., Paranjothy, S. (2014). Does the HPV vaccination programme have implications for cervical screening programmes in the UK? *Vaccine, 32* (16). 1828-1833. http://www.sciencedirect.com/science/article/pii/S0264410X14001509

[19] Congenital Anomaly Register and Information Service (CARIS). Public Health Wales. How CARIS works. http://www.caris.wales.nhs.uk/how-caris-works

[20] Charlton, R.A., Neville, A.J., Jordan, S., Pierini, A., Damase-Michel, C., Klungsøyr, K., Andersen, A.N., Hansen, A.V., Gini, R., Bos, J.H.J., Puccini, A., Hurault-Delarue, C., Brooks, C.J., de Jong-van den Berg, L.T.W., & de Vries, C.S. (2014). Healthcare databases in Europe for studying medicine use and safety during pregnancy. *Pharmacoepidemiology and Drug Safety, 23 (6),* 586-594. http://onlinelibrary.wiley.com/doi/10.1002/pds.3613/full