Millennium Cohort Study

A Guide to the NHS PEDW (inpatient and day case) Linked Administrative Data Sets: ICD-10 codes in Continuous Spells

Efrosini Setakis, and Emla Fitzsimons

User guide to the data (First Edition)

December 2017

This document is available in alternative formats.
Please contact the Centre for Longitudinal Studies.
tel: +44 (0)20 7612 6875
email: clsfeedback@ucl.ac.uk

Table of Contents

Table of tables and figures

Acknowledgements

1   Background

This data note describes the process of deriving variables from Patient Episode Database for Wales (PEDW), which was linked to cohort members from the Millennium Cohort Study following consent given at MCS4 (2008) for administrative health records to be accessed.

1.1 Background to the Millennium Cohort Study

The Millennium Cohort Study (MCS) is the fourth of Britain's world-renowned national longitudinal birth cohort studies. Each follows a large sample of individuals born over a limited period of time through the course of their lives, charting the effects of events and circumstances in early life on outcomes and achievements later on. They help show how trajectories of health, wealth, education, family and employment are interwoven for individuals and vary between them.

The MCS offers large-scale information about the new century's children, and the families who are bringing them up, for the four countries of the United Kingdom. Its first sweep, carried out during 2001-2002, laid the foundations for a major new longitudinal research resource, taking a new "year-long" cohort of around 19,000 babies. In England and Wales they were born over the 12 months starting in September 2000, and over 13½ months from late November 2000 in Scotland and Northern Ireland. Information was collected from parents when the babies were aged nine months. The sample design allowed for disproportionate representation of families living in areas of child poverty, and in areas of England with high ethnic minority populations.

The first sweep was followed up with subsequent sweeps in 2003/4 (when cohort members were at age 3). At the first sweep, data was collected from across the four countries in the UK, with 18,553 families across the UK; 11,533 sampled in England, 2,761 in Wales, 2,336 in Scotland, and 1,923 in Northern Ireland. Following sweeps took place in 2006/7 (age 5), 2008 (age 7), 2012 (age 11), with the age 14 survey due to take place in 2015. At the 4th sweep, during which consent for health data linkage was gained, 11,533 families were sampled in England, 2,761 sampled in Wales, 2,336 sampled in Scotland, and 1,995 sampled in Northern Ireland.

Details of the data available from the main survey are available from the CLS website and in particular the MCS Guide to the Datasets at www.cls.ioe.ac.uk/mcs.

1.2 The health record linkage process.

Health records collected by NHS Wales are held in the NHS Patient Episode Database for Wales (PEDW). This data set records all episodes of inpatient, outpatient, and day case activity in NHS hospitals in Wales for those consented MCS cohort members who were hospitalised in Wales.

At the MCS4 (2008) sweep, household members with 'parental responsibility' for cohort member(s) gave permission for 1880 of a total of 1995 cohort members to have their health records accessed for the purpose of linkage and subsequent research[1]. Of these 1880 cohort members, 1,879 were successfully linked, and 1,307 (69.9% of consented cohort members, 65.4% of the total cohort present in Wales at sweep 4) had PEDW inpatient records successfully extracted, i.e. had inpatient health records available to link. Table 1 shows the attrition of the Welsh sample at the consent and data extract stages.

Linkage of hospital identifiers to CLS cohort members, and subsequent extraction of cohort members' PEDW records was carried out by the SAIL Databank at the University of Swansea. The matching took the form of a process through which unique and personal identifiers attached to the cohort member and the hospital record were deemed to be the same. Once this matching had taken part, the record could be extracted. These extracted records were then anonymised and transferred securely to CLS.

Table 1: Comparing Welsh cohort sample size with number given consent to link and PEDW linked sample size.

	N	% retained of previous stage	% retained of cohort at MCS4	Attrition since MCS4 sweep
Welsh cohort size at MCS4 sweep	1,995			
Number given consent to link	1,880	94.2	94.2	115
Number successfully linked to an NHS number	1,877	99.9	94.2	118
Number present in PEDW inpatient data extract	1,306	69.6	65.4	689

The size of the Welsh MCS sample of cohort members at the 2008 sweep was 1,995. The number of cohort members who were given consent for their health records to be accessed was 1,880, an attrition of 115. Of these 1,877 were successfully linked to an NHS number providing a linkage rate of 99.8%. For these, health records from the PEDW inpatient and day case records were extracted for 1,306 cohort members, 69.6% of the linked cohort. These records exclude maternity records, therefore the remaining 30.4% of the linked cohort were either used hospitals outside Wales or have not been hospitalised since birth.

---

[1] Original MCS sample size in Wales was 2,799 babies

2   Linkage outcome

Table 2 shows the proportion of MCS cohort members for which PEDW inpatient and day case records were successfully extracted. These include

Table 2. Linkage Outcome

	Ever in Wales prior to 2012
Number of cohort members	1995
Number of consents for linkage gained (and cases sent for matching)	1,880
Number of consented participants who were successfully linked to an NHS number	1,877
Number of cases successfully matched and records extracted	1,306
% of cases successfully matched (as a proportion of cases sent for matched to an NHS number)	99.9%
% of cases with available inpatient data (as a proportion of consents gained)	69.6%

3   Creation of Derived Variables

In-hospital data were supplied structured using four files described in figure 1. Each hospital stay is a called spell that can be divided into a number of episodes. An episode is defined as the period a patient is under the care of one consultant within a provider spell. A provider spell is the continuous period of time that an admitted patient (using a bed) spends in the care of one NHS health care provider (Local Health Board provider in Wales). Spells are initiated with an admission episode and will end by one of three outcomes:

- discharge,

- transfer to another NHS provider (hospital within a different Local Health Board),

- death.

In approximately 90% of cases, the provider spell consists of just one consultant episode, in other cases it consists of more than one episode.[2]

3.1   Examining diagnoses within the cohort

All diagnoses in hospital records are recorded using ICD-10 codes. ICD-10 is the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list from the World Health Organisation (WHO). The ICD-10 codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases. The ICD-10 is used by the NHS to monitor the incidence and prevalence of disease and other health problems[3].

ICD-10 codes are divided broadly into twenty two categories, as shown in Table 4.

The categories shown here are broad, and each category contains a number of sub-categories. Because of the relatively small sample size of the extracted PEDW records, for the purposes of the data analysis conducted here these more refined categories will not be explored. This small sample size also presents a risk to cohort members' confidentiality, a small sample size increases the likelihood of a person being identified by the data (e.g. diagnosis code), and this is particularly a problem in the event of rare conditions.

This analysis explores the ICD-10 diagnosis codes. ICD-10 is the 10th revision of the International Statistical Classification of Diseases and Related Health Problems. ICD-10 codes are assigned to hospital patients to indicate diagnoses assigned to them during their time in hospital. The analysis focused upon in this report examines the diagnoses assigned to cohort members while admitted to hospital throughout their lifetime.

We identified the incidence of broad diagnostic categories in the Welsh MCS cohort, and the pattern of these diagnoses over time, from birth to age 11.

---

[2] NHS Wales Informatics Service – Information & Statistics (2013). Annual PEDW Data Tables – Notes & Definitions.
http://www.infoandstats.wales.nhs.uk/documents/869/NotesDefinitions_Oct131.pdf

[3] http://www.who.int/classifications/icd/en/
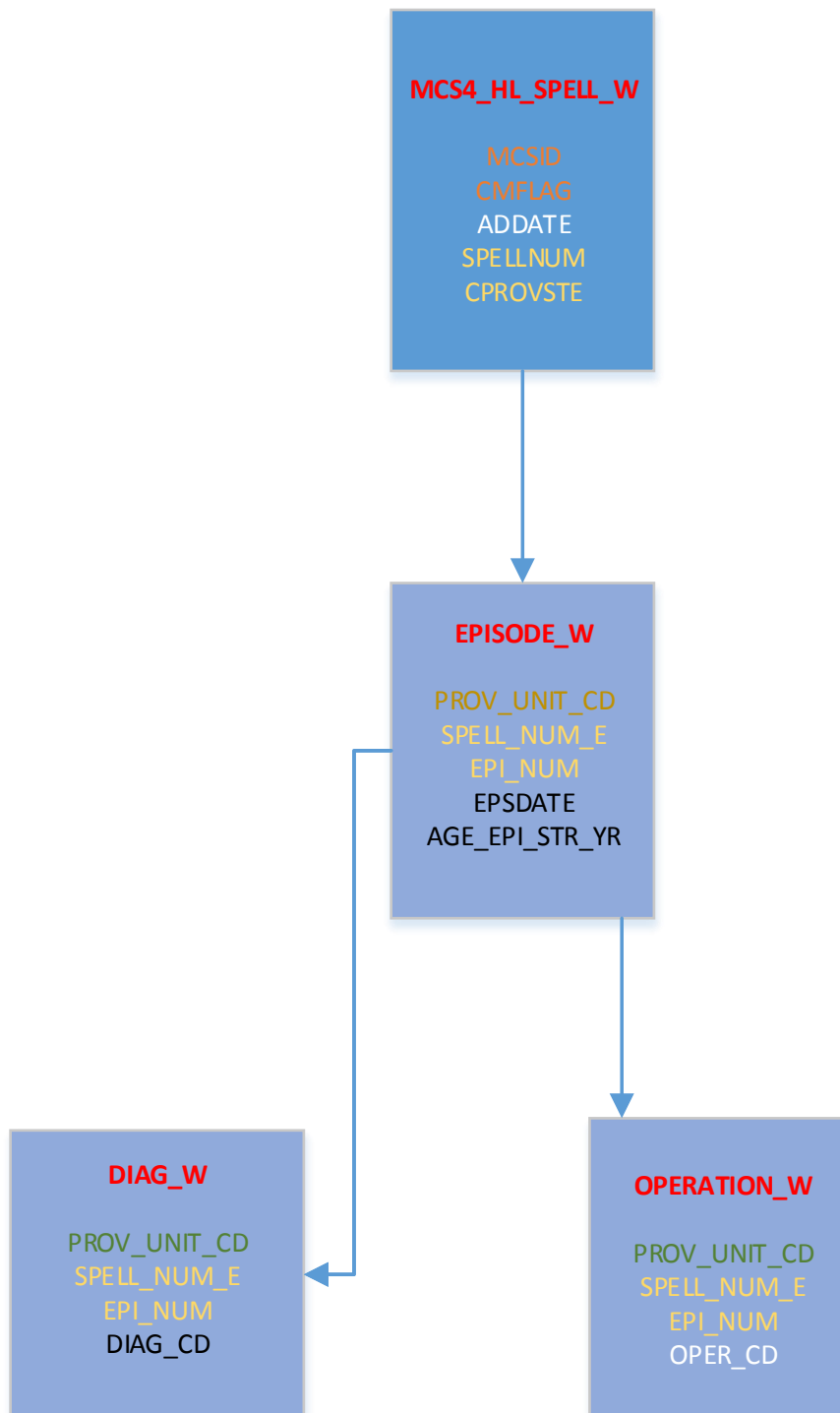
Figure 1: Structure of in-hospital data provided with linking variables and variables relevant to our analysis.

Table 4: ICD-10 codes and their descriptions, and number of cases represented in the MCS cohort member extract.

Chapter	Blocks	Title
1	A00–B99	Certain infectious and parasitic diseases
2	C00–D48	Neoplasms
3	D50–D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4	E00–E90	Endocrine, nutritional and metabolic diseases
5	F00–F99	Mental and behavioural disorders
6	G00–G99	Diseases of the nervous system
7	H00–H59	Diseases of the eye and adnexa
8	H60–H95	Diseases of the ear and mastoid process
9	I00–I99	Diseases of the circulatory system
10	J00–J99	Diseases of the respiratory system
11	K00–K93	Diseases of the digestive system
12	L00–L99	Diseases of the skin and subcutaneous tissue

13	M00–M99	Diseases of the musculoskeletal system and connective tissue
14	N00–N99	Diseases of the genitourinary system
15	O00–O99	Pregnancy, childbirth and the puerperium
16	P00–P96	Certain conditions originating in the perinatal period
17	Q00–Q99	Congenital malformations, deformations and chromosomal abnormalities
18	R00–R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
19	S00–T98	Injury, poisoning and certain other consequences of external causes
20	V01–Y98	External causes of morbidity and mortality
21	Z00–Z99	Factors influencing health status and contact with health services
22	U00–U99	Codes for special purposes

3.2   Incidence of ICD-10 codes

This analysis focuses on identifying the unique number of ICD-10 codes used within a hospitalisation defined as a spell in PEDW records. Duplicate records were removed, further if an ICD10 code was reported multi times even in separate episode within the spell they would only be accounted for once. Subsequently the ICD-10 codes where categories using the chapters described in table 4.

In table 5 we display the most frequent ICD-10 chapters found in the medical records of the Welsh MCS cohort members, from birth to age 11.

3.3   Aggregated diagnoses data table.

We created a summary table (described in Table 6) in order to release hospital data under Secure Access without jeopardising our participants' anonymity. We aggregated diagnostic codes by ICD-10 chapter and age year of diagnosis. Age year of diagnosis is defined by PEDW as the age of the patient on the first day of the episode in which the diagnosis was recorded.

Further we calculated the number of unique hospitalisations (spells) each child had in each age year. Spells were assigned the age of the patient in the first episode of that spell.

Finally we created an indicator variable wales_health_link to flag which participants were successfully linked to an NHS number. If a participant has been linked but there is no evidence of any hospitalisations this could mean that either they have been hospitalised outside Wales or these participants have not been hospitalised since birth.

Table 5: Incidence of ICD-10 codes in the PEDW extracted data for MCS cohort members, sorted by most common.

Chapter	Blocks	Title	Unique events within spells	% of total diagnostic codes
10	J00–J99	Diseases of the respiratory system	738	15.91
21	Z00–Z99	Factors influencing health status and contact with health services	592	12.76
18	R00–R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	565	12.18
1	A00–B99	Certain infectious and parasitic diseases	487	10.5
22	U00–U99	Codes for special purposes	354	7.63
19	S00–T98	Injury, poisoning and certain other consequences of external causes	327	7.05
20	V01–Y98	External causes of morbidity and mortality	301	6.49
11	K00–K93	Diseases of the digestive system	300	6.47
16	P00–P96	Certain conditions originating in the perinatal period	200	4.31
8	H60–H95	Diseases of the ear and mastoid process	136	2.93
17	Q00–Q99	Congenital malformations, deformations and chromosomal abnormalities	130	2.8

14	N00–N99	Diseases of the genitourinary system	116	2.5
12	L00–L99	Diseases of the skin and subcutaneous tissue	108	2.33
7	H00–H59	Diseases of the eye and adnexa	61	1.32
13	M00–M99	Diseases of the musculoskeletal system and connective tissue	57	1.23
6	G00–G99	Diseases of the nervous system	42	0.91
4	E00–E90	Endocrine, nutritional and metabolic diseases	39	0.84
3	D50–D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	32	0.69
9	I00–I99	Diseases of the circulatory system	20	0.43
5	F00–F99	Mental and behavioural disorders	18	0.39
2	C00–D48	Neoplasms	15	0.32
15	O00–O99	Pregnancy, childbirth and the puerperium	0	0

Table 6: Variables in aggregated ICD-10 dataset (Wales)

File name:

Variable	Description
MCSID	MCS Research ID
CMFLAG	CMFLAG
HEALTHF4	MCS4 Health Consent
wales_health_link	0/1 Indicator variable whether the participant was successfully linked to Welsh health records or not.
age	Age band for age of cohort member at diagnosis (as of episode admission date)
nhosp	Number of hospitalisations in that age year
icd_1	Number of diagnosis from ICD 10 chapter 1: Certain infectious and parasitic diseases
icd_2	Number of diagnosis from ICD 10 chapter 2: Neoplasms
icd_3	Number of diagnosis from ICD 10 chapter 3: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
icd_4	Number of diagnosis from ICD 10 chapter 4: Endocrine, nutritional and metabolic diseases
icd_5	Number of diagnosis from ICD 10 chapter 5: Mental and behavioural disorders
icd_6	Number of diagnosis from ICD 10 chapter 6: Diseases of the nervous system
icd_7	Number of diagnosis from ICD 10 chapter 7: Diseases of the eye and adnexa

icd_8	Number of diagnosis from ICD 10 chapter 8: Diseases of the ear and mastoid process
icd_9	Number of diagnosis from ICD 10 chapter 9: Diseases of the circulatory system
icd_10	Number of diagnosis from ICD 10 chapter 10: Diseases of the respiratory system
icd_11	Number of diagnosis from ICD 10 chapter 11: Diseases of the digestive system
icd_12	Number of diagnosis from ICD 10 chapter 12: Diseases of the skin and subcutaneous tissue
icd_13	Number of diagnosis from ICD 10 chapter 13: Diseases of the musculoskeletal system and connective tissue
icd_14	Number of diagnosis from ICD 10 chapter 14: Diseases of the genitourinary system
icd_15	Number of diagnosis from ICD 10 chapter 15: Pregnancy, childbirth and the puerperium
icd_16	Number of diagnosis from ICD 10 chapter 16: Certain conditions originating in the perinatal period
icd_17	Number of diagnosis from ICD 10 chapter 17: Congenital malformations, deformations and chromosomal abnormalities
icd_18	Number of diagnosis from ICD 10 chapter 18: Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
icd_19	Number of diagnosis from ICD 10 chapter 19: Injury, poisoning and certain other consequences of external causes
icd_20	Number of diagnosis from ICD 10 chapter 20: External causes of morbidity and mortality
icd_21	Number of diagnosis from ICD 10 chapter 21: Factors influencing health status and contact with health services

icd_22	Number of diagnosis from ICD 10 chapter 22: Codes for special purposes