Millennium Cohort Study Linked health administrative datasets - Scottish Medical Records (SMR)

User guide

First edition, October 2020

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

Contact

Questions and feedback about this user guide should be sent to clsfeedback@ucl.ac.uk.

How to cite this guide

Kerry- Barnard, S., Dennison, K., Fitzsimons, E. (2020) *Millennium Cohort Study: A guide to the linked health administrative datasets – Scottish Medical Records (SMR)*. London: UCL Centre for Longitudinal Studies.

Acknowledgements

CLS would like to thank NHS Digital for the provision of the linked data.

This guide was first published in October 2020 by the UCL Centre for Longitudinal Studies.

UCL Social Research Institute

University College London

20 Bedford Way

London WC1H 0AL

www.cls.ucl.ac.uk

The UCL Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Social Research Institute, University College London. It manages four internationally-renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit www.cls.ucl.ac.uk.

This document is available in alternative formats. Please contact the Centre for Longitudinal Studies.

tel: +44 (0)20 7612 6875

email: clsfeedback@uck.ac.uk

Contents

About the Millennium Cohort Study

The Millennium Cohort Study (MCS) is a longitudinal birth cohort study, following a nationally representative sample of approximately 19,000 people born in the UK at the turn of the century.

Through the study, we have captured rich information about the different aspects of cohort members' lives, from birth to childhood and adolescence, and we are continuing to keep up with them now they are adults.

As a multidisciplinary study, MCS is used by researchers working in a wide range of fields. Findings from MCS have influenced policy at the highest level, and today the study remains a vital source of evidence on the major issues affecting young people's lives.

1. Introduction

This user guide describes the linkage of medical records, child health reviews and prescription and immunisation records in Scotland (henceforth “Scottish Medical Records”) to cohort members’ data in the Millennium Cohort Study (MCS). These medical records have been made available to CLS on the basis of informed consent by the Information Services Division (ISD) part of NHS National Services Scotland. The records were linked to cohort members from the Millennium Cohort Study whose parents had given their consent for linkage to administrative health record data. The main aim of this data linkage is to enhance the research potential of the information we collect in the surveys, making it an even richer resource for the research community.

The MCS has had seven sweeps to date: at ages 9 months (MCS1), 3, 5, 7, 11, 14 and 17 (MCS7), collecting data from mothers and their resident partners, and cohort members, through home-based surveys.

At the MCS4 (2008) sweep, during which consent for health data linkage up to the age of 14 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.

2. Consent to health data linkage

At the MCS4 (2008) sweep, permission was sought from a parent or guardian to collect information about their child from other sources including routine records on education, medical and other health related records.

A leaflet about collecting information from other sources was included in the advance mailing to families. It explained in detail what information the survey wanted to collect, from which sources, who the survey wanted to collect the information from, and why.

During the interview, the parent or guardian completed a consent form which included a section for parental permission to release information from health records from birth up to their child's 14th birthday.

Consent was also sought from the cohort members themselves at age 17. Cohort members who refused consent to data linkage have not been included in the linked datasets, regardless of parental consent status.

Detailed information on the fieldwork and consent collection can be found in the MCS4 Technical Report and User Guide. All documents can be found under 'documentation' at https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/mcs-age-7-sweep/ .

3. Health data linkage

3.1 SMR datasets

The Scottish Medical Records (SMRs) are held by the Information Services Division (ISD) part of NHS National Services Scotland.. Together, they contain information about all hospital admissions in Scotland. It is comprised of seven data collections: Outpatient attendance (SMR00), Inpatient and day care attendance (SMR01), Scottish cancer registry (SMR06), Birth records/ neonatal records (SMR11), Scottish Immunisation & Recall System (SIRS), Prescribing Information System (PIS) and Child Health Reviews (CHR). The Child Health Reviews cover multiple reviews from birth to school records. The ISD website contains detailed information about each dataset, including background information, data dictionaries and definitions.[1]

Table 1: List of datasets provided

Dataset	Contents
SMR00	Outpatient attendance, years 2001-2015
SMR01	Inpatient and day care attendance, years 2000-2015
SMR06	Scottish cancer registry (linked but not shared due to sparse data)
SMR11	Scottish birth records /neonatal records, years 2000-2002
SIRS	Scottish Immunisation & Recall System, years 2000-2015
PIS	Prescribing Information System, years 2009-2015
CHR	Child Health Reviews from the health visitors, from birth to school age, years 2000-2015

[1] https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets/; https://www.isdscotland.org/Health-Topics/Prescribing-and-Medicines/; https://www.isdscotland.org/Health-Topics/Child-Health/Child-Health-Programme/

The linked Scottish health data cover diverse topics including: diagnosis, maternity, mortality, types of therapies, treatment length, Scottish Index of Multiple Deprivation (SIMD), Carstairs deprivation index, developmental milestones and anthropometry.

In 2015, CLS sought consent from ISD for linkage of SMR records for consented cohort members.

3.2 Matching strategy

Those MCS cohort members who resided in Scotland at any of the first five sweeps ("ever-Scottish") and whose parent or guardian gave consent to linkage of administrative medical records at sweep four were selected for linkage - 1,494 of a total of 2370 cohort members, representing 63% of the Scottish sample.

In order to conduct the linkage, identifying information (surname and first initial, full date of birth, sex, address including postcode, and hospital reference number), was sent securely to the ISD with a CLS proxy ID. At the ISD, these identifiers were used to match to Community Health Index (CHI) number, which was in turn used to match to the health records.

The CHI is a population register used in Scotland for health care purposes. The CHI number uniquely identifies a person on the index. The health records were then returned securely by ISD to the CLS data management team, and contained the CLS proxy ID but without the above mentioned identifying information and CHI. The health records could then be linked to the MCS data.

3.3 Matching rates

Table 2 shows the number of successful matches of SMR records following data linkage. Twins and triplets have been excluded from this dataset due to disclosure concerns. Altogether, 1481 cohort members had at least one linked SMR.

Table 2. Consent and linkage

Number of ever-Scottish cohort members in Sweep 4	2370
Number with valid consent	1494
Total number with matched SMR data	1481
Linkage rate, as proportion of consented	99%

Table 3. Number of matched cases per dataset

Dataset		**Number of research participants**
SMR00		1160
SMR01		834
SMR11		320
SIRS		1460
PIS		1333
CHR	First visit	1178
	6-8 weeks	1117
	8-9 months	1115
	22-24 months	1102
	39-42 months	983
	Preschoolers	534
	School attendees	777

4. The research datasets

The administrative data includes:

SMR00 – Outpatient attendance

SMR01 – Inpatient and day care attendance

SMR06 – Scottish cancer registry (linked but not shared due to sparse data)

SMR11 – Scottish birth records /neonatal records

SIRS - Scottish Immunisation & Recall System

PIS – Prescribing Information System

CHR – Child Health Reviews from the health visitors, from birth to school age.

Each data collection needs to be applied for individually.

4.1 Licensing

The linked SMR data have been processed by CLS and supplied to the UK Data Service (UKDS) under Secure Access Licence. Applicants wishing to access this data need to establish the necessary agreement with the UKDS and abide by the terms and conditions of the UKDS Secure Access licence.

An additional condition of the licensing is that it is not permitted to link SMR data to MCS data that include any Scottish geographies lower than ‘Scotland’.

Access to the SMR linked data will only be provided via the UKDS. The data may only be accessed through a secure private network (Secure Lab), via the researcher’s own institutional desktop PC or at the Safe Room at the UK Data Archive. Applicants wishing to access this data, need to establish agreement with the UKDS and abide by the terms and conditions of the UKDS Secure Access licence. Before gaining access, researchers must make an application detailing the intended analysis and provide a justification as to why this data is requested.

To apply to access this data please visit the UKDS website here- http://ukdataservice.ac.uk/use-data/secure-lab/about.aspx.

4.2 Data documentation provided

Users are advised to use the SMR datasets in conjunction with the data dictionaries and documents provided by CLS available via UKDS, as follows:

Documentation file	**File name**
User guide	MCS_ScottishHealth_UserGuide_v1.pdf
ISD Guides	PHS-Deprivation-Guidance-version-3-4.pdf CHSP-PS-Clinical-Guidelines-2018-06-26-FINAL.pdf
SMR Crib Sheets	SMR00_CRIB_080518.pdf SMR01_CRIB_080518.pdf SMR01Long_stay_Geriatric_CRIB_2014_April.pdf PIS_fields_for_researcher_v5_eDRIS Guidance PRISMS-data-manual-2.4.pdf
Data collection forms	Pre-Hall4-6-8-Week.pdf Pre-Hall4-8-9-Month.pdf Pre-Hall4-22-24-Month.pdf Pre-Hall4-39-42-Month.pdf Pre-Hall4-First-Visit.pdf Pre-Hall4-Pre-School.pdf
British National Formulary	BNF chapters and subchapters.pdf
ICD-10 codes	ICD-10: International statistical classification of diseases and related health problems-V1-eng.pdf ICD-10: International statistical classification of diseases and related health problems-V2-eng.pdf ICD-10: International statistical classification of diseases and related health problems-V3-eng.pdf

Documentation file	File name
OCPCS-4 codes	OPCS48 Metadata File Description V1.0.pdf OPCS48 ToCE Analysis Nov 2016 V1.0.xlsx OPCS48 ToCE Specification V0.1.pdf

Acronyms

Users may find useful to become familiar with the following list of acronyms used in the data dictionary and data labels.

BNF: British National Formulary

Carstairs: Carstairs and Morris Index; a combined measurement of deprivation based on postcode sector. This does not include health as a contributing factor.

ISD: Information Services Division, part of NHS Scotland

ICD-10: International Classification of Disease; this is used to codify diagnostic data.

OPCS-4: Office of Population Censuses and Surveys Classification of Interventions and Procedures version 4; this codifies procedures, operations and interventions taking place in a clinical setting.

SIMD: Scottish Index of Multiple Deprivation

SMR: Scottish Medical Record

UKDS: UK Data Service

UKDS Data Dictionaries

The UKDS data dictionaries provide detailed information for each of the datasets. They include the variables names, format, labels or titles, positions in each dataset. They also provide information of the values included in each variable.

ISD Guides

Several datasets include deciles of deprivation related to the “data zone” – i.e. neighbourhood corresponding to a patient’s postcode of residence. Several values have been provided: Carstairs 2001 and 2011, and the Scottish Index of Multiple Deprivation (SIMD) 2004, 2006, 2009 and 2012. The order of deprivation changes in scores from 2009 onwards, so care must be taken to ensure the most and least categories are not used the wrong way round. The PHS Deprivation Guidance should be referred to, in order to ensure proper use.

The Child Health Surveillance Programme (CHSP-PS) is explained in the clinical guidelines provided.

SMR Crib Sheets

The crib sheets from the ISD are available as supplementary documents. They contain the lookups for the codes used in most of the SMR datasets. These have been applied to the data, however some values did not correspond to any look-up.

Data Collection Forms

The Child Health review datasets are fully labelled, but the data collection forms should help understand the nature of the data collection process. This particularly applies to where there may be gaps in the data where only “yes” answers are recorded.

British National Formulary

Section, chapter and subchapter names corresponding to the drugs groups in the prescribing data can be found here. These can be cross referenced with the variable BNF subchapter (D_PIBNFPPC) in the PIS dataset.

International Classification of Disease v10 (ICD-10)

These supplementary files originate from the WHO website[2] and will only made available for approved projects:

- ICD-10: International statistical classification of diseases and related health problems-V1-eng.pdf
- ICD-10: International statistical classification of diseases and related health problems-V2-eng.pdf
- ICD-10: International statistical classification of diseases and related health problems-V3-eng.pdf

Researchers should refer to “ICD-10: International statistical classification of diseases and related health problems V1” to interpret the diagnostic codes in the APC and OP datasets, V2 and V3 may be of help in building lists of codes to search for by diagnosis.

OPCS4 Interventions and Procedures Classification System

To interpret the OPCS data, researchers need to use the following supplementary files[3]:

- OPCS48 ToCE Analysis Nov 2016 V1.0
- OPCS48 ToCE Specification V0.1
- OPCS48 Metadata File Description V1.0

4.3 Identifiers

MCSID and CNUM combine to form the anonymised unique cohort member identifier which is used to maintain the confidentiality of cohort members in the linked health records. The MCSID and CNUM can also be used to merge this data and other deposited Next Steps datasets.

[2] International statistical classification of diseases and related health problems, 10th revision, Fifth edition, 2016 https://apps.who.int/iris/handle/10665/246208, Accessed 24th August 2020

[3] The OPCS Classification of Interventions and Procedures, codes, terms and text is Crown copyright (2019) published by Health and Social Care Information Centre, also known as NHS Digital and licenced under the Open Government Licence available at www.nationalarchives.gov.uk/doc/open-government-licence/open-government-licence.htm.

4.4 Data processing

Variable names

Variables that have been altered, either by truncation, top coding, recoding or creation of a pseudo-anonymised key are named with the prefix D. For example the diagnosis variable diag_01 becomes D_diag_01 as it has been truncated to 3 characters.

Variable labels and value labels

Variable labels and value labels have been taken from the online Scottish Medical Records data dictionaries. Derived variables are labelled with “DV: [Variable name], [Truncated/Pseudonymised/Topcoded]”.

Exclusion of records based on consent

In line with the conditions of consent, only records from events occurring before the cohort member turned 14 could be included in the dataset.

In the SMR00 dataset, which included age, only those records occurring when the variable for age is less than or equal to 13 years were included.

The SMR01, PIS and CHR did not include age, so it needed to be derived. School age data, month and year of birth, and month and year at the time of the event were used to derive approximate age. Only records up to but excluding the month the cohort member turned 14 have been included, as it could not be determined at which point in the month the data were collected. In the case of SMR01, discharge date was used to avoid including episodes that began within the consent period but finished outside of it.

4.5. Data de-identification

CLS is committed to protect research participant's rights and avoid data disclosure and re-identification of individuals using one or more variables in the dataset or in combination with other existing data.

Dates of birth, small geographical details and rare cases that could easily lead to data disclosure have been removed to comply with the NHS National Services Scotland Statistical Disclosure Control Protocol.

Similarly, some variables had categories grouped into wider categories to avoid the possibility of data disclosure. Geographical data for those living in sparsely populated areas have been grouped. ICD-10 codes have been truncated. These variables are prefixed with "D_". Cohort members from families with multiple births have not been included.

The SMR datasets include clinical information regarding diagnoses and procedures. Diagnoses are coded using the International Classification of Diseases version 10 (ICD-10). The codes have been truncated to the first three characters as a disclosure control measure. Operations and other procedures have been classified using the OPCS Classification of Operations and Procedures, version 4 (OPCS-4). This has been truncated to chapter as a disclosure risk control.

The Public Health and Intelligence Caldicott Guardian for National Services Scotland has given permission for the data as described to be deposited with the UK Data Service (UKDS) for access via its Secure Lab.

The UKDS perform a certain level of disclosure control on the outputs generated by researchers in the Secure Lab. This involves looking at outputs with cells <10 and the process is outlined in their SDC Handbook which can be downloaded from: https://securedatagroup.org/sdc-handbook/

Health Board variables, which are based on geographical area, have been replaced with a random key, this key is consistent across the datasets. Other changes to datasets SMR00, SMR01 and SMR11 have been outlined in the appendices.

4.6 SMR00 – Outpatient Attendance

This data includes records for appointments in medical consultant outpatient clinics, meeting a consultant or senior member of their team outside of an outpatient clinic session and attendances to see a nurse or allied health professional who has legal and clinical responsibility for the patient. Where there are more than one specialties of the clinic, multiple records will be made unless there is a single “main specialty” of that clinic. Care should be taken by researchers to avoid double counting these multiple records.

See also http://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets//SMR00-Outpatient-Attendance/

4.7 SMR01 – General/Acute Inpatient and Day Case

These data refer to patients receiving care in the general/acute specialties and who are inpatients/ day cases admitted to NHS hospitals from non NHS locations or to NHS beds in non-NHS institutions.

SMR01 records are generated when:

1. Inpatients and day cases are admitted to NHS hospitals from locations external to the NHS.
2. Inpatients and day cases are admitted to contracted NHS beds in non-NHS institutions.
3. Inpatients and day cases change specialty (with or without a change of consultant) when:
    - An Inpatient transfers to become an Inpatient in another specialty in the same hospital.
    - An inpatient becomes a Day Case in another specialty during the inpatient stay.
    - A day case transfers to become an Inpatient in another specialty (except when the day case episode is during an inpatient stay and the patient is transferring back to the original consultant).
    - A day case transfers back to resume an Inpatient stay but does NOT transfer back to the original consultant for medical reasons.

4. Inpatients and day cases transfer from another NHS hospital (including contracted NHS beds in non-NHS institutions).
5. Inpatients and day cases change consultant for medical reasons within the same specialty when :
    - An inpatient transfers as an Inpatient to the care of a different consultant for medical reasons in the same specialty in the same hospital.
    - An inpatient becomes a Day Case in the same specialty under a different consultant for medical reasons during the inpatient stay.
    - A day case transfers to become an Inpatient under the care of a different consultant in the same specialty for medical reasons (except when the day case episode is during an inpatient stay and the patient is transferring back to the original consultant).
    - A day case transfers back to resume an Inpatient stay in the same specialty but does NOT transfer back to the original consultant for medical reasons.
6. Inpatients move into and/or out of one of the valid significant facilities.
7. Inpatients return to hospital having been on pass for more than 5 days.

A new record is generated for a change of specialty or a hospital transfer or their inpatient/day case status is changed; this is indicated in the variable MANAGEMENT_OF_PATIENT. Therefore one stay may have more than one episode.

See also: http://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets//SMR01-General-Acute-Inpatient-and-Day-Case/

4.8 SMR11 – Scottish Birth Record/ Neonatal records

This has been superseded by the SBR in 2003 and ceased being updated in 2011. This still covers the period of birth for all of the cohorts. Despite the name this contains information on infants who require specialist or other care after birth, and contains some birth information but mostly clinical indicators. SMR11 records were created for babies with a congenital anomaly or who require medical care other than resuscitation.

SMR11 records were generated when these babies were:

- admitted to an NHS hospital
- admitted to a contracted NHS bed in a non-NHS institution
- admitted following delivery
- transferred from another NHS hospital
- changed consultants for medical reasons within the same specialty
- moved into or out of one of the significant facilities
- admitted as private patients to NHS or private facilities

As a result there may be multiple records for a single stay in hospital.

See also: https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Appendices/Appendix-G/Metadata/index.asp?Search=S&ID=720&Title=SMR11 - Neonatal - Discontinued (SMR Metadata)

4.9 SIRS - Scottish Immunisation & Recall System

Contains the records of childhood immunisation that took place within the NHS. Immunisations taking place outside of the NHS may not be present. This is in wide format – there is only one record per cohort member. This covers vaccines for:

Diptheria – a course of 3 injections, plus a booster

Tetanus – a course of 3 injections, plus a booster

Pertussis - a course of 3 injections, plus a booster

Polio – three doses, plus a booster and a follow up dose

Haemophillus Influenza B, three doses, plus a booster and a catch up round

MeningococcalC, three doses, ~~a teenage dose~~ and an extra course

Pneumococcal (PCV),catch up dose only

Measles, Mumps and Rubella (MMR), initial dose and booster

BCG (Tuberculosis vaccine), single dose

Measles, single dose (not MMR), given twice.

Mumps, single dose (not MMR)

Rubella, single dose (not MMR)

H1N1 virus, two doses

Hepatitis B, three doses plus a booster

These additional vaccines were also included however they fell after the cohort member's fourteenth birthday and could not be shared

Human Papilloma Virus (HPV), three doses given to girls.

School leaving booster TdIPV (Tetanus, diphtheria and polio), given once

Annual influenza vaccine 2013-2015, one dose per year

See also: https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?ID=4&SubID=12

4.10 PIS – Prescribing Information System

The PIS contains prescription and dispensary information linked to cohort member: including all prescriptions dispensed in Scotland outside of a hospital setting and those written in Scotland but dispensed elsewhere in the UK but excluding those dispensed in hospital but including those written in hospital but dispensed elsewhere. Each record reflects an individual dispensary episode.

This dataset uses the British National Formulary (BNF) code for each prescribed item that has been dispensed. The first two characters of the code refer to the chapter of the BNF, the second two characters refer to the section and the third two characters refer to the paragraph; the family of drugs. A copy of the section, chapter and subchapter names is available in the supplementary files.

See also: http://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=9

4.11 CHR - Child Health Review

These are the reviews from the health visitors, who record information about the child's progress routinely. Between 2000 and 2006 these were:

First visit (about 10 days after the birth of the child)

- 6-8 weeks visit
- 8-9 months visit
- 21-24 months visit
- 39-42 months visit
- Pre-school review
- Repeated school reviews

The health visitor records include information about breastfeeding, environmental factors such as parental smoking, immunisation and health and development indicators. Parental concern variables are repeated throughout, but these are collected as an affirmative tick box, therefore the answer "no" is incomplete, and it is therefore difficult to distinguish missing values from negative ones.

There may be multiple records for a cohort member for the repeated school reviews; these can be differentiated using the date that the height and weight were measured (variable name DATE_HW_CHSC). For the other health review records there is a maximum of one record per cohort member, as each review takes place only once.

The Data Collection Instruments used to collect Child Health Visitor information are described in the table below.

Visit	Data Collection Form(s) & Notes	N	Years collected
First visit	• Current form https://isdscotland.scot.nhs.uk/Health-Topics/Child-Health/Child-Health-Programme/First_Visit.pdf • 2000 form https://isdscotland.scot.nhs.uk/Health-Topics/Child-Health/Child-Health-Programme/Pre-Hall4-First_Visit.pdf • Pre 2015 form https://isdscotland.scot.nhs.uk/Health-Topics/Child-Health/Child-Health-Programme/_docs/First-Visit-up-to-2015.pdf • The form used for data collection is not indicated in the dataset.	1178	0-88 days old
6-8 week visit	• 2000 form https://www.isdscotland.org/Health-Topics/Child-Health/Child-Health-Programme/Pre-Hall4-6-8-Week.pdf • Pre 2015 form https://www.isdscotland.org/Health-Topics/Child-Health/Child-Health-Programme/docs/6-8-Week-up-to-2015.pdf. • The form used for data collection is not indicated in the dataset.	1138	2001-2003
8-9 month visit	• https://www.isdscotland.org/Health-Topics/Child-Health/Child-Health-Programme/Pre-Hall4-8-9-Month.pdf	1135	2001-2002
22-24 month visit	• https://www.isdscotland.org/Health-Topics/Child-Health/Child-Health-Programme/Pre-Hall4-22-24-Month.pdf	1122	2002-2004
39-42 month visit	• https://www.isdscotland.org/Health-Topics/Child-Health/Child-Health-Programme/Pre-Hall4-39-42-Month.pdf	997	2002-2006

Visit	Data Collection Form(s) & Notes	N	Years collected
Preschool review	https://www.isdscotland.org/Health-Topics/Child-Health/Child-Health-Programme/Pre-Hall4-Pre-School.pdf.	545	No date information
School review	https://www.isdscotland.org/Health-Topics/Child-Health/Child-Health-Programme/Pre-Hall4-P1-screening.pdf . Unlike the other reviews which target one age group, there are multiple records per cohort member (CM)	1109 / 790 unique CMs	2005-2016

5. Disclosure control: UKDS requirements for data users

As the HES data linked to the longitudinal NCDS data are only available via the UKDS Secure Lab, the UK Data Service will always perform a certain level of disclosure control on the outputs generated by researchers, as outlined in their SDC Handbook, which can be downloaded from https://securedatagroup.org/sdc-handbook/.

The two UK Data Service Secure Lab rules of thumb that will be applied to all outputs are:

- Threshold rule: No cells should contain less than 10 observations;
- Dominance rule: No observation should dominate the data to a huge extent.

Appendix 1. Modifications to the Inpatient data (SMR00)

Variable name	SMR original variable name	Variable description	Modification
D_PATIENT_CATEGORY	PATIENT_CATEGORY	Patient Category	Rare values recoded to missing
D_MAIN_OPERATION_A	MAIN_OPERATION_A	Main Operation A	Truncated to 1 char
D_MAIN_OPERATION_B	MAIN_OPERATION_B	Main Operation B	Truncated to 1 char
D_OTHER_OPERATION_1_A	OTHER_OPERATION_1_A	Other Operation 1 A	Truncated to 1 char
D_OTHER_OPERATION_2_A	OTHER_OPERATION_2_A	Other Operation 2 A	Truncated to 1 char
D_HBRES_CURRENTDATE	HBRES_CURRENTDATE	Health Board of residence	Pseudonymised with new key
D_HBTREAT_CURRENTDATE	HBTREAT_CURRENTDATE	Health Board of treatment	Pseudonymised with new key

Appendix 2. Modifications to the Inpatient data (SMR01)

Variable name	SMR original variable name	Variable description	Modification
D_CIS_MARKER	CIS_MARKER	Continuous Inpatient Stay Marker	Top coded at 15 days
D_PATIENT_CATEGORY	PATIENT_CATEGORY	Patient Category	Rare values recoded to missing
D_MAIN_CONDITION	MAIN_CONDITION	Main Condition	Truncated to 3 chars
D_OTHER_CONDITION_1	OTHER_CONDTION_1	Other Condition 1	Truncated to 3 chars
D_OTHER_CONDITION_2	OTHER_CONDITION_2	Other Condition 2	Truncated to 3 chars
D_OTHER_CONDITION_3	OTHER_CONDITION_3	Other Condition 3	Truncated to 3 chars
D_OTHER_CONDITION_4	OTHER_CONDITION_4	Other Condition 4	Truncated to 3 chars
D_OTHER_CONDITION_5	OTHER_CONDITION_5	Other Condition 5	Truncated to 3 chars
D_MAIN_OPERATION_A	MAIN_OPERATION_A	Main Operation A	Truncated to 1 char
D_MAIN_OPERATION_B	MAIN_OPERATION_B	Main Operation B	Truncated to 1 char
D_OTHER_OPERATION_1_A	OTHER_OPERATION_1_A	Other Operation 1 A	Truncated to 1 char
D_OTHER_OPERATION_1_B	OTHER_OPERATION_1_B	Other Operation 1 B	Truncated to 1 char
D_OTHER_OPERATION_2_A	OTHER_OPERATION_2_A	Other Operation 2 A	Truncated to 1 char
D_OTHER_OPERATION_2_B	OTHER_OPERATION_2_B	Other Operation 2 B	Truncated to 1 char
D_OTHER_OPERATION_3_A	OTHER_OPERATION_3_A	Other Operation 3 A	Truncated to 1 char

Variable name	SMR original variable name	Variable description	Modification
D_OTHER_OPERATION_3_B	OTHER_OPERATION_3_B	Other Operation 3 B	Truncated to 1 char
D_HBRES_CURRENTDATE	HBRES_CURRENTDATE	Health Board of residence	Pseudonymised with new key
D_HBTREAT_CURRENTDATE	HBTREAT_CURRENTDATE	Health Board of treatment	Pseudonymised with new key

Appendix 3. Modifications to the Births data (SMR11)

Variable name	SMR original variable name	Variable description	Modification
D_MAIN_CONDITION	MAIN_CONDITION	Main Condition	Truncated to 3 chars
D_OTHER_CONDITION_1	OTHER_CONDTION_1	Other Condition 1	Truncated to 3 chars
D_OTHER_CONDITION_2	OTHER_CONDITION_2	Other Condition 2	Truncated to 3 chars
D_OTHER_CONDITION_3	OTHER_CONDITION_3	Other Condition 3	Truncated to 3 chars
D_OTHER_CONDITION_4	OTHER_CONDITION_4	Other Condition 4	Truncated to 3 chars
D_OTHER_CONDITION_5	OTHER_CONDITION_5	Other Condition 5	Truncated to 3 chars
D_OTHER_CONDITION_6	OTHER_CONDITION_6	Other Condition 6	Truncated to 3 chars
D_OTHER_CONDITION_7	OTHER_CONDITION_7	Other Condition 7	Truncated to 3 chars
D_MAIN_OPERATION_A	MAIN_OPERATION_A	Main Operation A	Truncated to 1 char
D_MAIN_OPERATION_B	MAIN_OPERATION_B	Main Operation B	Truncated to 1 char
D_OTHER_OPERATION_1_A	OTHER_OPERATION_1_A	Other Operation 1 A	Truncated to 1 char
D_ESTIMATED_GESTATION	ESTIMATED_GESTATION	Estimated length of gestation in weeks	Banded values lower than 31 weeks
D_BIRTH_WEIGHT	BIRTH_WEIGHT	Birthweight in grams	Banded values under 2 kg
D_DISCHARGE_WEIGHT	DISCHARGE_WEIGHT	Discharge weight in grams	Banded values under 2 kg

Variable name	SMR original variable name	Variable description	Modification
D_NEO_NATAL_STAY	NEO_NATAL_STAY	Length of neonatal stay	Top coded at 30 days
D_LENGTH_OF_STAY	LENGTH_OF_STAY	Length of stay	Top coded at 30 days
D_ HB_OF_RESIDENCE	HB_OF_RESIDENCE	Health Board of residence	Pseudonymised with new key
D_HB_OF_TREATMENT	HB_OF_TREATMENT	Health Board of treatment	Pseudonymised with new key