

Primary Care Working Group Datasets Project
June 2017

'Datasets that may be of interest to Primary Care Researchers in the UK'

Dr Lucy McDonnell, Professor Brendan Delaney, Professor Frank Sullivan

On behalf of The Farr Institute UK

www.farrinstitute.org

Table of Contents

1	INTRODUCTION	4
2	ELECTRONIC MEDICAL RECORD DATA	5
2.1	NATIONAL	5
2.1.1	<i>Clinical Practice Research Datalink</i>	5
2.1.2	<i>Q Research</i>	7
2.1.3	<i>THIN Database</i>	9
2.1.4	<i>RCGP Research and Surveillance Centre</i>	11
2.2	REGIONAL	13
2.2.1	<i>Research One (England)</i>	13
2.2.2	<i>SAIL databank: Primary care GP Dataset (Wales)</i>	15
2.2.3	<i>Primary Care Clinical Informatics Unit Research (PCCIU) (Scotland)</i>	17
2.3	LOCAL	19
2.3.1	<i>Lambeth DataNet (LDN)</i>	19
2.3.2	<i>Clinical Effectiveness Group East London Database</i>	21
3	QUALITY OF PRIMARY CARE SERVICES	23
3.1	QUALITY AND OUTCOMES FRAMEWORK (QOF) DATA	23
4	PRESCRIBING DATA	25
4.1	PRIMARY CARE PRESCRIBING DATA FOR ENGLAND	25
4.2	PRIMARY CARE PRESCRIBING DATA FOR WALES	27
4.3	PRIMARY CARE PRESCRIBING DATA FOR SCOTLAND	29
4.4	PRIMARY CARE PRESCRIBING DATA FOR NORTHERN IRELAND	30
5	AUDIT	32
5.1	NATIONAL AUDIT OF CANCER DIAGNOSIS IN PRIMARY CARE	32
5.2	NATIONAL DIABETES AUDIT (NDA)	33
6	HEALTH SURVEYS	34
6.1	GENERAL PRACTICE PATIENT SURVEY	34
6.2	NATIONAL HEALTH SURVEYS	36
6.2.1	<i>Health Survey for England</i>	36
6.2.2	<i>Scottish Health Survey (SHeS)</i>	37

6.2.3	<i>Welsh Health Survey (WHS)</i>	39
6.2.4	<i>Health Survey Northern Ireland</i>	41
6.3	CANCER PATIENT EXPERIENCE SURVEYS	42
6.4	SCOTTISH DIABETES SURVEYS	44
7	SPECIAL DATASETS	45
7.1	ABERDEEN MATERNITY AND NEONATAL DATABANK	45
7.2	SCOTTISH HEALTH INFORMATICS CENTRE (HIC) SERVICES	47
8	COHORT STUDIES	49
8.1	NATIONAL	49
8.1.1	<i>UK Biobank</i>	49
8.1.2	<i>MRC National Survey for Health and Development (NSHD)</i>	51
8.1.3	<i>Boyd Orr Cohort</i>	53
8.1.4	<i>HALCyon (Healthy Ageing across the Life Course) cross-cohort collaboration</i>	54
8.2	REGIONAL	57
8.2.1	<i>Born in Bradford (England)</i>	57
8.2.2	<i>Aberdeen Birth Cohorts (Scotland)</i>	59
8.2.3	<i>Growing up in Scotland</i>	61
9	ADMINISTRATIVE DATASETS	63
9.1	SCOTTISH ADMINISTRATIVE RECORDS	63
10	SCREENING DATASETS	65
10.1	UK CANCER SCREENING	65
11	USEFUL WEBSITES	67

1 Introduction

The Primary Care Working Group datasets project was developed to fill a gap in the literature, regarding the existence and availability of healthcare datasets relevant to primary care researchers in the United Kingdom. Although established primary care researchers often have historic links to datasets, there are potentially many other users who are unaware of the existence of such a large variety of datasets, or have previously encountered barriers to access them.

The aims of this project were to systematically identify, describe and facilitate access to as wide a variety of healthcare datasets based in England, Scotland, Wales and Northern Ireland, as possible. We aimed to ensure that website addresses and contact details were up to date at the time of press. Custodians of major datasets were given the opportunity to verify the information before publication.

The details given are intended as a starting point for researchers and to increase awareness of what is available. We have included both major popular datasets as well as other, lesser well-known and often free resources for researchers. By giving references to peer-reviewed publications, we have illustrated the variety of research that can be carried out with routinely collected data and how the data has previously been used.

We hope this document will prove to be a useful resource for both established and new researchers in the field of primary care. The authors welcome contributions from other dataset custodians who may wish for their details to be included.

2 Electronic Medical Record Data

2.1 National

2.1.1 Clinical Practice Research Datalink

Name and Description	Type of Data	Coverage	Geography	Duration
<p>Clinical Practice Research Datalink (CPRD)</p> <p>Governmental, not-for-profit research service, jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare Products Regulatory Agency (MHRA), a part of the Department of Health</p>	Primary care	All patients registered at General Practices contained in the CPRD (subset of all practices in the UK), who have not dissented from secondary use of GP patient data	UK wide	Since 1987
Context of data collected	Volume	Events for data collected	Granularity	Coding
The data are usually entered directly by the GP, or if not, by other surgery staff, usually during the consultation with the patient present. In some cases, the data are entered after the appointment	693 contributing practices across the UK Covers 14.2 million patients of which 2.8 million are currently active	<ul style="list-style-type: none"> • GP registration • Patient appointment/consultation • Clinical events (medical diagnoses) • Referrals to specialists and secondary care • Diagnostic tests • Lifestyle information • Prescribing events 	Per registered patient	Read codes
Extraction/Data collection	Consent	Access		
Primary care data are available online via CPRD GOLD. CPRD GOLD provides powerful disease and drug coding dictionaries and a fast query tool that 'defines' patient cohorts. An 'extract' tool then enables, as specified, cuts of data against a cohort or control group Data feeds are established with the GP systems providers (feeds from INPS Vision, EMIS and SystemOne –TPP)	Patient opt out GP practice opt-in	<p>https://www.cprd.com/intro.asp Primary care data are available online via CPRD GOLD</p> <p>Dependent on approval of a study protocol by the MHRA Independent Scientific Advisory Committee (ISAC)</p> <p>Data costs are charged at a fixed rate depending upon which data sources are required and complexity of linkage.</p> <p>An annual licence for online access to the CPRD GOLD database enables any number of individual ISAC-approved studies to be carried out.</p>		

Contact details	Linkage	Example Publications
<p>The knowledge centre Kc@cprd.com The clinical Practice Research Datalink Group The Medicines and Healthcare products Regulatory Agency 5th Floor 151 Buckingham Palace Road Victoria London SW1W 9SV England 020 3080 6383</p>	<p>Linked data are available for the following datasets</p> <ul style="list-style-type: none"> • Hospital Episode Statistics (HES) Admitted Patient Care (HES APC) • HES Outpatient (HES OP) data • HES Accident and Emergency (HES A&E) data • HES Diagnostic Imaging Dataset (HES DID) • Death Registration data from the Office for National Statistics • Cancer Registration data from Public Health England (PHE) • Cardiovascular disease registry data from the Myocardial Ischaemia National Audit Project (MINAP) • Measures of relative deprivation at Lower Layer Super Output Area (LSOA) level, based on patient and practice postcode 	<ul style="list-style-type: none"> • Menamin UCM et al., <i>Metformin use and survival from lung cancer: A population-based cohort study</i> Lung Cancer 2016 Apr 01;94:35-39 • Clarke AT. et al., <i>High dose atorvastatin associated with increased risk of significant hepatotoxicity in comparison to simvastatin in UK GPRD Cohort</i> PLoS One 2016; 11(3): e0151587 • Charlton R. et al., <i>Asthma medication prescribing before, during and after pregnancy: a study in seven European regions</i> BMJ Open 2016; 6(1): e009237 • Kontopantelis E. et al., <i>Primary care consultation rates among people with and without severe mental illness: a UK cohort study using the Clinical Practice Research Datalink</i> BMJ Open 2015 Dec 16;5(12):e008650

2.1.2 Q Research

Name and Description	Type of Data	Coverage	Geography	Duration
<p>QRESEARCH is a large consolidated database derived from the pseudonymised health records of over 22 million patients</p> <p>Non-profit making collaboration between The University of Nottingham and EMIS</p>	Primary Care	<p>Patients registered at select General Practices which use the EMIS system</p> <p>Currently registered or historical patients</p> <p>Covers a population of 22 million</p>	UK wide	<p>Historical records extend back to early 1990s</p> <p>Established in 2003</p>
Context of data collected	Volume	Events for data collected	Granularity	Coding
The data are usually entered by the GP during or after the consultation	Approximately 1200 GP practices: over 18 million patients	<ul style="list-style-type: none"> • Year of birth • Sex • Registration dates • Date of death • Deprivation score • Ethnicity • Consultations • Prescriptions • Clinical values e.g. blood pressure • Test results • Clinical events/diagnoses • Only coded data are collected, no free text or strong patient identifiers 	Per registered patient	<p>Read codes and Snomed CT codes for clinical diagnoses;</p> <p>EMIS own drug dictionary for medication</p>
Extraction/Data collection	Consent	Access		
<p>GP data are collected during routine practice and are pseudonymised, uploaded and transmitted via the N3 network to a dedicated server in EMIS</p> <p>EMIS uploads key census related variables such as deprivation and maker of rurality into the patient's electronic record</p>	Patient opt out GP practice opt-in	<p>www.qresearch.org</p> <p>Researchers need to define a research question or hypothesis, then contact QRESEARCH to determine if it's data are suitable</p> <p>Data are available for use in answering pre-defined research questions only</p> <p>Most suitable for case control studies, cohort, sample size calculations</p> <p>Funding needs securing before access to data are granted</p>		

<p>Hospital Episode Statistics data are supplied by the Health and Social Care Information Centre Mortality data are supplied by the Office of National Statistics Cancer registry data are supplied by Public Health England All data from all four sources are pseudonymised at the source before transmission to the University of Nottingham using the www.openpseudonymiser.org tool.</p>		<p>The licence agreement will be between QRESERACH and a UK university. At least one member of the research team must be medically qualified registered with the GMC Researchers will be provided with anonymised patient level data required to answer research question – not whole dataset Research is expected to be published in peer-reviewed journals</p>
<p>Contact details</p>	<p>Linkage</p>	<p>Example Publications</p>
<p>Professor Julia Hippisley-Cox Julia.hippisley-cox@nottingham.ac.uk</p>	<p>QRESEARCH database is linked to other sources of data such as cancer registries, death data and hospital episode statistics on an ongoing basis. Certain linked data are only available on-site at the University of Nottingham though projects can be done in collaboration with external researchers</p>	<ul style="list-style-type: none"> • Vinogradova Y, Coupland C, Hippisley-Cox J <i>Use of combined oral contraceptives and risk of venous thromboembolism: nested case-control studies using the QResearch and CPRD databases.</i> BMJ 2015; 10.1136/bmj.h2135 [Available online] • Coupland CA, Hill T, Morriss R, Arthur A, Moore M, Hippisley-Cox J. <i>Antidepressant use and risk of suicide and attempted suicide/self-harm in people aged 20 to 64: cohort study using a primary care database.</i> BMJ 2015;350:h517 [Available online] • Hippisley-Cox J, Coupland CA. <i>Predicting risk of upper gastrointestinal bleed and intracranial bleeding with anticoagulants: prospective cohort study to derive and validate the QBleed score.</i> BMJ 2014; 349; g4606 [Available online]

2.1.3 THIN Database

Name and Description	Type of Data	Coverage	Geography	Duration
<p>UK primary care electronic medical record (EMR) data resource</p> <p>Patients are representative of the UK population by age, gender, medical conditions and death rates adjusted for demographics and social deprivation</p> <p>GPs contributing to THIN provide health services under the terms of the UK's National Health Service (NHS)</p>	Primary Care	Patients registered at General practices contributing to THIN database. 5.67% coverage of UK in 2013	UK wide	Since 1987
Context of data collected	Volume	Events for data collected	Granularity	Coding
The data are usually entered by the GP during or after the appointment	Active GP practices = >395 Active patients = >3.1M Total patients = >15M	<ul style="list-style-type: none"> Demographics (including registration details) Diagnoses (including symptoms, tests, results, referrals) Prescribing Additional Health Information e.g. smoking, alcohol intake Free text comments Socioeconomic data from THIN (based on postcode) 	Per registered patient	Read code Unified version 2
Extraction/Data collection	Consent	Access		
<p>Practices which use Vision software and have joined THIN are provided with unobtrusive anonymous data collection software</p> <p>Upon joining THIN an initial Full Data Collection (FDC), which includes all retrospective data are sent to THIN</p> <p>Following this incremental data are collected</p>	Patient opt out GP practice opt-in	<p>www.epic-uk.org</p> <ul style="list-style-type: none"> Sublicense: enables access to entire period (minimum 1 year) to conduct unlimited studies, subject to protocol approval Data extracts: subsets of raw data can be provided in accordance with researchers' study protocols and specifications Summaries and reports can be provided following detailed study 		

<p>automatically and downloaded electronically each month</p>		<p>specifications. Summaries can be provided in MS Excel spreadsheets or SAS datasets</p> <p>Advice is offered to researchers on the feasibility of conducting studies using THIN including preliminary information for studies e.g. approx. frequencies of diseases or prescriptions</p>
<p>Contact details</p>	<p>Linkage</p>	<p>Example Publications</p>
<p>hbhullar@uk.imshealth.com + 44 203 075 5675</p>	<p>THIN data can be linked with Hospital Episodes Statistics (HES) for use within Europe</p>	<ul style="list-style-type: none"> • Szatkowski L and Aveyard P. Provision of smoking cessation support in UK primary care: impact of the 2012 QOF revision Br J Gen Pract 2016 Jan;66(642)e10-5 • Lycett, D. et al., <i>The association between smoking cessation and glycaemic control in patients with type 2 diabetes: a THIN database cohort study</i>, Lancet Diabetes Endocrinol 2015 3(6):423-30 • Boursi B et al: <i>The effect of past antibiotic exposure of diabetes risk</i>, Eur K Endocrinol 2015 172(6):639-48

2.1.4 RCGP Research and Surveillance Centre

Name and Description	Type of Data	Coverage	Geography	Duration
<p>Part of RCGP Clinical Innovation and Research Centre (CIRC). Active research and surveillance unit collecting all coded data (including prescription data), from practices in England. The Centre is England's primary source for Influenza surveillance.</p> <p>RSC publishes 2 weekly reports:</p> <ul style="list-style-type: none"> • Weekly returns service: a report which includes weekly incidence rates of over 30 communicable and respiratory diseases; • Gastroenteritis report: which includes weekly incidence rates of intestinal infectious diseases and pathogens associated with gastroenteritis <p>The core surveillance and vaccine effectiveness work is mainly funded by Public Health England.</p>	Primary Care data	Approx. 1.5% of English population	England	Established in 1957
Context of data collected	Volume	Events for data collected	Granularity	
<p>Patient data are entered by the GP or administrative staff during or after patient consultations.</p> <p>Some practices supply microbiology samples (nasal/throat swabs)</p> <p>Data providing practices receive individual feedback on data quality, based on coding.</p>	<p>230 GP Practices in RSC network; data currently collected from 160 practices</p> <p>Approx. 1,500,000 patients</p>	<p>Patient demographics</p> <p>Diagnoses</p> <p>Risk factors for chronic conditions</p> <p>Medication prescribing</p> <p>Vaccine uptake</p>	Per patient	
Extraction/Data collection	Consent	Access	Coding	
Data are extracted by an initial bulk upload when practice joins the network; incremental uploads take	GP practice opt in Patient opt out	http://www.rcgp.org.uk/clinical-and-research/our-programmes/research-and-surveillance-centre.aspx	Read coding, SNOMED-coded data are also extracted (currently not analysed).	

<p>place weekly to create the RSC's reports uploads. Ten years' worth of historic data are stored and managed by the University of Surrey, which act as the data and analysis hub for the RCGP RSC. Data are pseudonymised as close to source as possible, encrypted for transmission to Surrey. The secure network at University of Surrey meets the NHS information security standard for holding these data as set out in the Information Governance toolkit. All patient coded data are extracted (including deregistered and deceased patients).</p>		<p>All planned research studies subject to ethical approval</p>	
<p>Contact details</p>	<p>Linkage</p>	<p>Example Publications</p>	
<p>i.yonova@surrey.ac.uk mariya.hriskova@rcgp.org.uk</p>	<p>RCGP RSC data can be linked to other types of datasets, including HES, using a hashing algorithm for NHS numbers</p>	<ul style="list-style-type: none"> • de Lusignan S, et al. RCGP Research and Surveillance Centre Annual Report 2014-2015: disparities in presentations to primary care. Br J Gen Pract. 2017 Jan;67(654):e29-e40. • Correa A, et al. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. BMJ Open. 2016 Apr 20;6(4):e011092. 	

2.2 Regional

2.2.1 Research One (England)

Name and Description	Type of Data	Coverage	Geography	Duration
<p>Health and Care research database developed by TPP (clinical software) in partnership with university of Leeds and the UK Government's Technology Strategy Board</p> <p>ResearchOne data comes from electronic patient records and the clinical system, SystmOne</p> <p>Not for profit</p>	<p>Primary Care (and Secondary care)</p>	<p>5000 providers including 2200 general practices, 1250 district nursing practices, 120 out of hours practices, 50 Child health units, 50 minor injuries/A&E units</p>	<p>England only</p>	<p>ResearchOne launched in 2013 SystmOne records date from before this time (varies)</p>
Context of data collected	Volume	Events for data collected	Granularity	Coding
<p>Information entered into electronic patient record by GP or other healthcare provider</p>	<p>30 million patients in whole dataset</p> <p>18 million primary care GMS registered patients (GP dataset)</p>	<ul style="list-style-type: none"> • Diagnoses • Procedures • Pathology test data • Prescribing data • Deprivation indices • Care pathways 	<p>Per registered patient</p>	<p>Primary care - CT3V (Clinical Terms Version 3 (Read codes))</p> <p>ResearchOne can support mapping of different coding schemes</p>
Extraction/Data collection	Consent	Access		
<p>Database consists of de-identified clinical and administrative data drawn from electronic patient records currently held on TPP SystmOne clinical system from a variety of organisations</p> <p>Database updated on a weekly basis</p>	<p>Patient opt out GP practice opt in</p>	<p>www.researchone.org</p> <p>Principal investigator should submit an expression of interest to the ResearchOne team with a brief summary of the project (form available on website)</p> <p>Pricing based on size and complexity of the required data</p>		

<p>Data are centrally checked on a continuous basis to assess data integrity, quality and representation</p> <p>Data are checked against national prescribing and mortality rates, aggregate census data, deprivation and rurality indices</p>		
<p>Contact details</p>	<p>Linkage</p>	<p>Example Publications</p>
<p>Research@tpp-uk.com 0113 205 0082</p>	<p>SystemOne records integrate data from multiple organisations into one record per patient. The research records contain linked data from wide variety of settings across primary and secondary care e.g. General Practice, Child Health, Community Health, Palliative Hospital, Out-of-hours Accident & Emergency, and Acute Hospitals</p>	<p>Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. <i>Age Ageing</i>. 2016 May;45(3):353-60.</p> <ul style="list-style-type: none"> • Powrie D.J. et al., <i>P282 An Integrated IT system for COPD between primary, secondary and community care using SystemOne</i>, <i>BMJ Thorax</i> 2012; 67:A188-A189 • Moore P. et al., <i>Medicines reconciliation using a shared electronic health care record</i>, <i>J Patient Saf</i> 2011 7(3):148-54 <p>Current projects:</p> <ul style="list-style-type: none"> • Mother and Child Outcomes from the Whooping Cough Vaccination Programme • Impact of text messages on vaccination uptake • Calcium supplementation and cardiovascular risk • ASPIRE: supporting evidence based practice • Improving prevention of cardiovascular events in primary care

2.2.2 SAIL databank: Primary care GP Dataset (Wales)

Name and Description	Type of Data	Coverage	Geography	Duration
<p>Secure Anonymised Information Linkage</p> <p>SAIL is an anonymised data linkage system which brings together routinely collected data for research, data and evaluation</p> <p>GP Data (Primary Care GP dataset) forms part of the SAIL dataset</p>	Primary Care	Welsh practices that have signed up to SAIL (currently 73% of Welsh Practices)	Wales	From January 2000 to present (varies by practice)
Context of data collected	Volume	Events for data collected	Granularity	Coding
<p>GP clinical data</p> <p>The majority of the data are entered by the clinician during the patient consultation</p>	Around 2.25 million patients (75% of total Welsh population)	<ul style="list-style-type: none"> • data capture of signs, symptoms • test results • diagnoses • prescribed treatment • referrals for specialist treatment • social information 	Per registered patient	Mainly READ codes but sometimes local codes used
Extraction/Data collection	Consent	Access		
<p>Data are securely extracted from the clinical system using the SAIL module in Audit+</p> <p>Test results are electronically transferred from secondary care systems</p>	GP Practice opt in Patient opt out	<p>www.saildatabank.com</p> <p>Initial contact via form on SAIL website to request scoping discussion or via SAILdatabank@swansea.ac.uk</p> <p>Information Governance Review Panel (IGRP) Application form to be completed</p> <p>All users are required to complete safe researcher training before access to SAIL data granted</p>		
Contact details	Linkage	Example Publications		
SAILdatabank@swansea.ac.uk	SAIL is involved with many organisations that provide anonymous version of datasets to the SAIL databank	<ul style="list-style-type: none"> • Farroha A et al., <i>Using anonymised, routinely collected health data in Wales to estimate the incidence of depression after burn injury</i>, J Burn Care Res 2013 34(6):644-8 • Pickrell et al., <i>Weight change associated with antiepileptic drugs</i>, J 		

	<p>The core datasets are:</p> <ul style="list-style-type: none"> Annual District Birth Extract Annual District Death Extract Bowel Screening Wales (BSW) Cervical Screening Wales (CSW) Congenital Anomaly Register and Information Service (CARIS) Emergency department Data Set (EDDS) National Community Child Health Database (NCCHD) Outpatient Dataset (OPD) Patient Episode Database for Wales (PEDW) Primary Care GP dataset UK Health Dimensions Welsh Cancer Intelligence and Surveillance Unit (WCISU) Welsh Demographic Service (WDS) 	<p>Neurol Neurosurg Psychiatry 2013 84(7):796-9</p> <ul style="list-style-type: none"> • Brophy S et al., <i>No increased rate of acute myocardial infarction or stroke among patients with ankylosing spondylitis – a retrospective cohort study using routine data</i>, Semin Arthritis Rheum 2012 42(2):140-5
--	--	---

2.2.3 Primary Care Clinical Informatics Unit Research (PCCIU) (Scotland)

Name and Description	Type of Data	Coverage	Geography	Duration
<p>PCCIU was founded in 1999 to provide a reporting service to Scottish General Practice in collaboration with RCGP Scotland</p> <p>It is the first Scottish Primary care dataset</p> <p>PCCIU collected practice data for the SPICEpc (Scottish Programme for Improving Clinical Effectiveness in Primary Care) Programme bi-annually in spring and autumn between 2000 and 2011. The collected data for each season was amalgamated to a single seasonal general practice database</p>	Primary Care	GP Practices in Scotland that agreed to participate in the data collection process Approx. 15% of the Scottish population	Scotland	Data collected biannually from 2000 until 2011 Subsets of the data available for researchers
Context of data collected	Volume	Events for data collected	Granularity	Coding
<p>GP clinical data</p> <p>Data was entered by the GP during or after the consultation, or by administrative staff</p> <p>Prescription data and diagnostic coding for clinical summaries is expected to be quite complete</p> <p>Items related to QOF were recorded from 2003</p>	<p>The number of practices contributing data to a given time period varies as the data were collected from practices during different seasons and years</p> <p>In 2006, 393 practices contributed (2,311,937 patients)</p> <p>Spring 2007 – 393</p>	<ul style="list-style-type: none"> • Patient encounters • Diagnoses, results, problems and procedures • Details of measurements • Details of issued prescriptions • Date of birth (month/year), gender and postcode to sector level • Registered clinician • Clinician cypher 	Per registered patient	Read codes

	<p>practices Autumn 2007 – 336 Spring 2008- 299 Autumn 2008 -267 Spring 2009 -244 Autumn 2009- 220 Spring 2010 -194 Autumn 2010 -139 Spring 2011- 87</p>			
Extraction/Data collection	Consent	Access		
Data was extracted from practice electronic records bi-annually from 2000 to 2011 using electronic software	GP practice opt in	http://www.abdn.ac.uk/iahs/research/primary-care/pcciu/access.php Researchers should email first to discuss the research proposal and see whether the database is suitable Following that a research protocol should be submitted according to the submission process It normally takes between 3-4 months from submission of protocol until the data are released for analysis Costs apply		
Contact details	Linkage	Example Publications		
pcciureserach@abdn.ac.uk Dr Chris Burton (Research manager)	Not available	<ul style="list-style-type: none"> • Agur K et al. <i>How does sex influence multimorbidity? Secondary analysis of a large nationally representative dataset.</i> International Journal of Environmental Research and Public Health, 13(4), p391 • Macfarlane TV et al., <i>Aspirin and other non-steroidal anti-inflammatory drug prescriptions and survival after the diagnosis of head and neck and oesophageal cancer</i> Cancer Epidemiol 2015 Dec;39 (6):1015-22 • Barnett K et al;, <i>Epidemiology of multimorbidity and implications for health care, research and medical education. A cross-sectional study.</i> Lancet 2012, 380, 37-43 		

2.3 Local

2.3.1 Lambeth DataNet (LDN)

Name and Description	Type of Data	Coverage	Geography	Duration
<p>Lambeth DataNet uses GP patient records to obtain anonymous information to help plan and improve healthcare services</p> <p>DataNet is managed by NHS Lambeth Clinical Commissioning group who work with partners from Lambeth and Southwark Public Health Teams, King's College London and Lambeth GPs.</p>	Primary Care	Patients registered at General Practices contributing to DataNet who have not dissented from secondary use of GP data	London Borough of Lambeth, England	Records dating back to 2008 can be analysed
Context of data collected	Volume	Events for data collected	Granularity	Coding
Socio-demographic data are usually entered by administrative staff. Clinical data are usually entered by the GP during or after the appointment.	<p>Active GP practices = 47</p> <p>Active patients = approx. 350,000</p>	<ul style="list-style-type: none"> • Socio -demographics including residential area, age, gender, ethnicity, language preference, country of birth and religion • Clinical information: diagnoses including long term conditions e.g. diabetes, hypertension, cholesterol levels, test results • Prescribing and medication 	Per registered patient	Read codes
Extraction/Data collection	Consent	Access		
<p>Information comes from forms patients completed when they register with a GP practice and their treatment records. Information is anonymised before it leaves the practice.</p> <p>Data are extracted centrally using EMIS Web on an annual basis</p>	<p>Patient opt out</p> <p>GP practice opt in</p>	<p>http://www.lambethccg.nhs.uk/your-health/Information-for-patients/Pages/DataNet.aspx</p> <p>Potential researchers should be affiliated with King's College London, the Biomedical Research Centre at Guys and St Thomas' NHS Foundation Trust, Lambeth CCG or Lambeth Public Health.</p> <p>Research questions should first be discussed with Dr Mark Ashworth</p>		

Contact details	Linkage	Example Publications
<p>Dr Mark Ashworth markashworth@nhs.net</p>	<p>In some cases, data can be linked with hospital or Accident and Emergency department records. For some studies, particularly those where larger sample sizes are needed or data on Bangladeshi/South Asian patients LDN has pooled data with the East London Database (no data linkage, no special permission needed beyond R&D approval) (<i>see Mathur paper</i>) LDN has been linked to CRIS (Case Register Interactive Research) database held at the Institute of Psychiatry)(<i>see Woodhead paper</i>)</p>	<ul style="list-style-type: none"> • Woodhead C, et al., <i>Patterns of physical co-/multimorbidity among patients with serious mental illness: a London borough-based cross sectional study</i> BMC Fam Pract 2014 Jun 11; 15:117 • Mathur R. et al., <i>Ethnicity and stroke risk in patients with atrial fibrillation</i> Heart 2013 Aug; 99 (15):1087-92 • Schofield P. et al., <i>Managing hypertension in general practice: a cross-sectional study of treatment and ethnicity</i> Br J Gen Pract 2012 Oct;62(603):e703-9 • Pinto R. et al., <i>Differences in the primary care management of patients with psychosis from two ethnic groups: a population-based cross-sectional study</i> Fam Pract 2010 Aug;27(4):439-46

2.3.2 Clinical Effectiveness Group East London Database

Centre for Primary Care and Public Health. Queen Mary University of London.

Name and Description	Type of Data	Coverage	Geography	Duration
<p>The CEG East London database uses GP electronic records to obtain non-identifiable patient information which is to help plan and improve healthcare services</p> <p>CEG East London data is managed by the east London Clinical Effectiveness Group which works with partner organisations City & Hackney, Newham and Tower Hamlets CCGs and Local Authority public health teams, QMUL/UCLP researchers, the local GP provider Federations, Barts Health Trust and the Commissioning Support Unit.</p>	Primary Care	<p>Patients registered at General Practices in the three CCGs City & Hackney, Newham and Tower Hamlets. Population approximately 1 million currently registered individuals.</p> <p>Data is available by</p> <ul style="list-style-type: none"> • Patient* • Lower Super Output Area • LA Ward • Practice • CCG <p>*Data will not be released at individual patient level without patient/controller consent. Cell size fewer than 5 and must be aggregated.</p> <p>Sensitive data such as HIV, abortions will require further consideration for governance purposes as will data at small area level.</p>	CCGs/London Boroughs of City and Hackney, Newham and Tower Hamlets	<p>EMIS records of currently registered dating back to the earliest date of their electronic registration back to approx. 1990.</p> <p>Patients who left the practice prior to 2008 are not accessible.</p>
Context of data collected	Volume	Events for data collected	Granularity	
<p>Self-reported ethnic group is available on around 90% of adults and Townsend /IMD scores (based on postcode/LSOA) on 99%.</p> <p>Clinical data is usually entered by GP clinical staff in</p>	<p>Active GP practices = 143/147</p> <p>Active patients = approx. 1,000,000</p>	<ul style="list-style-type: none"> • Socio -demographics including residential area, age, gender, ethnicity, language preference, country of birth and religion • Clinical information: 	<p>Per registered patient</p> <p>Though data for use should be aggregate and non-identifiable</p>	

association with a GP practice consultation identifiable as face to face, by telephone or by a third party.		<p>diagnoses including long term conditions e.g. diabetes, hypertension</p> <ul style="list-style-type: none"> • Diagnostic tests: e.g. Haemoglobin, cholesterol • Prescribing and medication • Some limited hospital attendance e.g. A&E 	
Extraction/Data collection	Consent	Access	Coding
<p>Data is extracted centrally from EMIS Web.</p> <p>Some data is held in an annually updated JSNA database and other information is extracted on a per request basis.</p>	<p>Patient opt out GP practice opt in</p> <p>All access to the data is subject to governance by CEG on behalf of the data controllers</p>	<p>http://www.blizard.qmul.ac.uk/ceg-home.html</p> <p>Potential researchers should be affiliated with a UK university. Research access should first be discussed with Dr John Robson Clinical lead CEG. Reader QMUL.</p>	<p>Read codes for EMIS. ICD10 and HRGs for SUS</p>
Contact details	Linkage	Example Publications	
<p>John Robson j.robson@qmul.ac.uk</p> <p>Room 2.13 Centre for Primary Care and Public Health Queen Mary University of London, Yvonne Carter Building. 58 Turner Street, London E1 2AB</p> <p>Admin: 020 7882 2553</p>	<p>There are options for linkage to SUS data from BartsHealth for registered patients in east London</p> <p>Linked data are only available from 2010</p>	<p>Schofield P, et al. <i>Does depression diagnosis and antidepressant prescribing vary by location? Analysis of ethnic density associations using a large primary-care dataset.</i> Psychol Med. 2016;46:1321-9.</p> <p>Homer K, et al. <i>Statin prescribing for primary prevention of cardiovascular disease: a cross-sectional, observational study.</i> Br J Gen Pract. 2015;65:e538-44.</p> <p>Robson J, et al. <i>The NHS Health Check programme: implementation in east London 2009-2011.</i> BMJ Open. 2015;5:e007578.</p>	

3 Quality of primary care services

3.1 Quality and Outcomes Framework (QOF) data

Name and Description	Type of Data	Coverage	Geography	Duration
<p>The QOF was introduced as part of the General Medical Services (GMS) contract on 1 April 2004</p> <p>QOF contains 3 or 4 domains</p> <p>Each domain consists of a set of achievement measures known as indicators against which practices score points according to level of achievement</p> <p>2014-5 QOF measured achievement against 81 indicators in England</p> <p>the QOF agreement differs between countries in the UK</p> <p>Practices achieve points for e.g. managing chronic disease, how well the practice is organised, improving quality and productivity in other parts of the health service and offering extra services</p>	Primary Care	<p>General Practices in England, Scotland, Wales and Northern Ireland</p> <p>Covers approximately 56 million registered patients in England</p>	UK wide (England, Scotland, Wales and N. Ireland each manage own data)	<p>Since 2004</p> <p>Most recent QOF data are from April 01 2014 to March 31 2015</p>
Context of data collected	Volume	Events for data collected	Granularity	Coding
<p>Individual clinical data are usually entered by the GP during or after the consultation</p> <p>GP practices then submit summary data per practice</p>	<p>ENGLAND: 2014-5 data are available for 7,779 general practices in England</p> <p>98.7% of practices opt in</p> <p>SCOTLAND: Virtually all GP practices in</p>	<p>Practices are measured against 81 indicators which cover 19 clinical areas (including disease prevalence) and 6 public health aspects of GP practice activity (may vary between countries)</p> <p>Details of indicators for England can be found at www.qof.hscic.gov.uk</p>	Per GP practice	Specific QOF indicators are used

	Scotland take part WALES: 459 GP practices took part 2014-5 N IRELAND: 350 GP practices took part 2014-5			
Extraction/Data collection	Consent	Access		
Primary use of QOF data are to support QOF payments to GPs ENGLAND: Data are extracted using the Calculating Quality Reporting Service (CQRS), together with the General Practice Extraction Service (GPES) SCOTLAND: Data are extracted from the QOF Calculator Database WALES: Data derived from national 'CM Web' software N IRELAND: Data are extracted from QMAS (Quality Management and Analysis System)	GP practice opt in	Under the Open Government Licence data can be accessed and used free of charge ENGLAND: Summary and practice level data available via www.qof.hscic.gov.uk SCOTLAND: Summary data and practice level data available via http://www.isdscotland.org/health-Topics/General-Practice/Quality-And-Outcomes-Framework/ WALES: Summary and practice level data available via http://gov.wales/statistics-and-research/general-medical-services-contract N. IRELAND: Summary level data available via https://www.health-ni.gov.uk/topics/doh-statistics-and-research/quality-outcomes-framework-qof		
Contact details	Linkage	Example Publications		
ENGLAND - via HSCIC - general email is enquiries@hscic.gov.uk SCOTLAND - nss.isdGeneralPractice@nhs.net WALES - stats.healthinfo@wales.gsi.gov.uk N IRELAND - qofdataenquiries@health-ni.gov.uk	It is not possible to directly compare the QOF performance of the four nations of the UK	<ul style="list-style-type: none"> McKay AJ. et al., <i>Are primary care factors associated with hospital episodes for adverse drug reactions? A national observational study</i> BMJ Open 2015 Dec 29;5(12):e008130 Martin JL. et al., <i>Physical health indicators in major mental illness: data from the Quality and Outcome Framework in the UK</i> Lancet 2015 Feb 26;385 Suppl 1:S61 		

4 Prescribing data

4.1 Primary Care prescribing data for England

Name and Description	Type of Data	Coverage	Geography	Duration
<p>Every month NHS England publishes anonymised data about the drugs prescribed by GPs Figures are based on information systems at NHS Prescription Services, part of the NHS Business Services Authority The figures are collected as part of the process of reimbursing dispensers for drugs supplied.</p> <p>The figures are based on prescriptions written in primary care (by general medical practitioners, and other primary care prescribers e.g. nurses, pharmacists, chiropodists) in England and dispensed anywhere in the UK</p>	<p>Primary care level prescribing data</p>	<p>NHS GP practices in England</p>	<p>England</p>	<p>Since 2010 CCG level data since 2013 Prior to April 2013 PCT level prescribing data was published</p>
Context of data collected	Volume	Events for data collected	Granularity	
<p><u>GP prescribing data</u> – includes prescribing by GPs and non-medical prescribers attached to practices <u>CCG prescribing data</u> – as above but also includes prescriptions which cannot be attributed to a CCG but are funded directly by an Area Team <u>PCT prescribing data</u> – as above but also includes prescribing which cannot be attributed to a specific practice</p>	<p>Number of practices vary from month to month as practices may close and new practices open, practices may merge or divide</p>	<p>General practice – list of all medicines, dressings and appliances that are prescribed and dispensed each month For each practice in England information is presented as follows;</p> <ul style="list-style-type: none"> • Total number of items prescribed and dispensed • Total net ingredient cost • Total actual cost • The total quantity 	<p>Per practice, per month PCT or CCG prescribing is quarterly</p>	

Extraction/Data collection	Consent	Access	Coding
<p>NHS Prescription Services, part of the NHS Business Services Authority (BSA) provides the data GP practice level prescribing data are obtained from the BSA's prescribing and dispensing information systems</p>	<p>Not required</p>	<p>www.openprescribing.net is a website designed by the University of Oxford to be a user friendly way to access the data at both practice and CCG/PCT level</p> <p>Alternatively, the data files can be accessed via www.hscic.gov.uk/gpprescribingdata – GP practice level prescribing can be downloaded but the file sizes are very large (over 1GB)</p> <p>For PCT/CCG level data there is a link to www.hscic.gov.uk/iview which is a free resource to view data for which researchers will need to register</p>	<p>Prescribing data are represented using BNF codes GP practices are identified by their national code Additional linked files provide practice codes, names and addresses, chemical names and BNF codes</p>
Contact details	Linkage	Example Publications	
<p>enquiries@hscic.gov.uk</p>	<p>Potentially prescribing information can be linked to other data that uses the practice code</p>	<ul style="list-style-type: none"> Ashworth M., et al., <i>Antibiotic prescribing and patient satisfaction in primary care in England: cross-sectional analysis of national patient survey data and prescribing data</i>. Br J Gen Pract 2016 Jan;66(642):e40-6 Rowlingson B et al., <i>Mapping English GP prescribing data: a tool for monitoring health-service inequalities</i> BMJ Open 2013 Jan 3;3(1) 	

4.2 Primary care prescribing data for Wales

Name and Description	Type of Data	Coverage	Geography	Duration
<p><u>General Practice Prescribing Data Extract</u> - data covers prescriptions that are prescribed in Wales by GPs and non-medical prescribers that have prescribed on behalf of the GP practice, that are then dispensed within Wales or England. Private prescriptions are not included.</p> <p><u>GP Practice analysis</u> – data are derived from prescriptions issued from GP Practices in Wales and dispensed by; Chemists in England and Wales Appliance contractors in England and Wales Dispensing Doctors in Wales</p> <p>Data shows the number of items prescribed by each practice by month and the number of patients registered at the each practice</p>	Primary care level prescribing data	NHS GP Practices in Wales	Wales	Since April 2013, practice analysis since 2010
Context of data collected	Volume	Events for data collected	Granularity	Coding
Data are published on the 1 st working day of each month and will be available for the current and two complete financial years.	Number of practices varies	<p><u>GP Prescribing data extract</u>: list of all medicines, dressings and appliances that are prescribed and dispensed each month</p> <p>For each practice in Wales information is presented as follows;</p> <ul style="list-style-type: none"> • Total number of items prescribed and dispensed • Total net ingredient cost • Total actual cost • The total quantity of drug prescribed • Defined Daily Dose (DDD) • ADQ – Assumed average maintenance dose per day for a drug used for its main indication in adults 	Per registered practice, per month	BNF codes and practice codes are used

		<ul style="list-style-type: none"> Year and month to which data relates 		
Extraction/Data collection	Consent	Access		
The information is obtained from prescribing and dispensing information systems	Not required	<p>GP Prescribing data extract: http://www.primarycareservices.wales.nhs.uk/general-practice-prescribing-data-extrac</p> <p>Data can be downloaded from the website</p> <p>Each compressed file contains three individual file which can be viewed in applications such as Microsoft WordPad</p> <p>GP Practice analysis: http://www.primarycareservices.wales.nhs.uk/gp-practice-analysis</p> <p>Data can be downloaded for each year in the form of excel files</p>		
Contact details				
Prescribing.management@wales.nhs.uk				

4.3 Primary Care prescribing data for Scotland

Name and Description	Type of Data	Coverage	Geography	Duration
Prescribing Information System (PIS) – a definitive data source for all prescribing relating to all medicines and their costs that are prescribed and dispensed in the community in Scotland.	Primary care level prescribing data	All medicines and their costs that are prescribed and dispensed in the community in Scotland	Scotland	From April 1993 onwards
Context of data collected	Volume	Events for data collected	Granularity	Coding
Includes prescriptions written in Scotland but dispensed anywhere in the UK Prescriptions written by GPs and other prescribers e.g. nurses and dentists Also included are prescriptions written in hospitals but dispensed in the community	Around 90 million data items are loaded per annum	<ul style="list-style-type: none"> • CHI number (Community Health Index) • Prescriber - location, type • Dispenser - location, type • Patient – age, gender • Geography/Organisation/Deprivation e.g. council area code • Prescription date, prescribable item • Prescription types, BNF code • Prescribed, dispensed or paid information 	Per practice, health board or national level, per annum	BNF drug codes
Extraction/Data collection	Consent	Access		
The data are generated as a by-product of prescription processing by NSS Practitioner services for the payment of dispensing contractors	Not required	www.adls.ac.uk Summary prescribing information system data are available by contacting nss.isdprescribing@nhs.net All applications for ISD non-patient identifiable data and patient identifiable data must be made through eDRIS (Electronic Data Research and Innovation Service) via NSS.eDRIS@nhs.net Charges apply		
Contact details	Linkage	Example Publications		
nss.isdprescribing@nhs.net nss.eDRIS@nhs.net	It is possible to related prescribing data to other ISD datasets holding the CHI number, from 2009 onwards	<ul style="list-style-type: none"> • Johnson C.F. et al., <i>Benzodiazepine and z-hypnotic prescribing for older people in primary care: a cross-sectional population based study</i> Br J Gen Pract 2016 Apr 25 • Sharpe K. et al., <i>Reduced risk of oestrogen receptor positive breast cancer among peri-and post menopausal women in Scotland following a striking decrease in use of hormonal replacement therapy</i> Eur J Cancer 2010 Mar; 46(5): 937-43 • Morrison J. et al., <i>Factors influencing variation in prescribing of antidepressants by general practices in Scotland</i> Br J Gen Pract 2009 Feb;59(559):e25-31 		

4.4 Primary Care prescribing data for Northern Ireland

Name and Description	Type of Data	Coverage	Geography	Duration
<p>Prescribing by GP Practice dataset covers prescriptions that are prescribed in Northern Ireland in primary care</p> <p>Data includes all prescribed medicines, dressings and appliances dispensed each month since 2013</p> <p>The GP prescribing database originates from the Enhanced Prescribing Database (EPES)</p> <p>From April 2016 the GP Prescribing Dataset will be updated on a monthly basis</p>	<p>Primary care level prescribing data</p>	<p>All NHS GP practices in Northern Ireland</p>	<p>Northern Ireland</p>	<p>Practice level data since April 2013</p>
Context of data collected	Volume	Events for data collected	Granularity	Coding
<p>The data covers prescription prescribed in Northern Ireland by GPs or other prescribers (attached to a GP practice) that are dispensed by a community pharmacist, dispensing doctor or appliance supplier, are submitted to the BSO for payment and have been paid</p>	<p>The number of GP practices varies from month to month as practices may close and new practices open, practices may merge or divide</p>	<p>For each practice and for each medicine (presentation), dressing and appliance the following information is provided</p> <ul style="list-style-type: none"> • Number of prescribed items that are dispensed • The quantity of tablets, capsules, liquid etc. dispensed • The gross cost and actual cost 	<p>Per practice per month</p>	<p>GP practice codes BNF Codes</p>
Extraction/Data collection	Consent	Access		
<p>The Business Services Organisation (BSO) provides the data. GP practice level prescribing data are obtained from the BSO's prescribing and dispensing information systems</p>	<p>Not required</p>	<p>http://www.hscbusiness.hscni.net/services/2471.htm</p> <p>Data can be downloaded from the website</p> <p>For every quarter, data for individual months are provided separately and data for each month consists of approximately 450,000 rows of data</p>		

Contact details	Linkage	Example Publications
Responsible Statistician: Mr. Sandy Fitzpatrick Telephone: 028 9053 2965 E-mail: info.bso@hscni.net	The data can be potentially linked to other published data that uses the practice code	<ul style="list-style-type: none"> • Barry H. et al., <i>Potentially inappropriate prescribing among people with Dementia in Primary Care: A retrospective cross sectional study using the enhanced prescribing database</i> J Alzheimers Dis 2016 Apr 11 • Sweeney J. et al., <i>Inappropriate prescribing of combination inhalers in Northern Ireland: retrospective cross-sectional cohort study of prescribing practice in primary care</i> Prim Care Resp J 2014 Mar;23(1):74-8

5 Audit

5.1 National audit of cancer diagnosis in primary care

Name and Description	Type of Data	Coverage	Geography	Duration
This audit was undertaken in 2009/10 as part of the National Awareness and Early Diagnosis Initiative. It was led by Professor Greg Rubin, with the RCGP, the National Cancer Action Team and the National Cancer Intelligence Network	Primary care audit	20 Cancer Networks, 14% of all practices in England	England	Conducted between April 2009 and April 2010
Context of data collected	Volume	Events for data collected	Granularity	
Clinical information is routinely collated in electronic patient records	Data were collected on 18,879 patients by 1170 practices in 20 cancer networks	<ul style="list-style-type: none"> • Stage of cancer at diagnosis • Number of times patient attended surgery • Investigations ordered • Symptoms at presentation • Intervals along the patient pathway 	Per patient	
Extraction/Data collection	Consent	Access		
Participating GP practices were required to complete the audit template from their practice clinical records and hospital correspondence	Practice agreed to participate	Summary report data are available from http://www.rcgp.org.uk/policy/rcgp-policy-areas/national-audit-of-cancer-diagnosis-in-primary-care.aspx		
Contact details	Example Publications			
Project lead: Professor Greg Rubin Greg.rubin@durham.ac.uk	<ul style="list-style-type: none"> • Keeble S et al., <i>Variation in promptness of presentation among 10,297 patients subsequently diagnosed with one of 18 cancers: evidence from a National audit of cancer diagnosis in primary care</i> Int J Cancer 2014 Sep 1; 135(5):1220-8 • Lyratzopoulos G et al., <i>Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patient with 18 common and rarer cancers</i> Br J Cancer 2013 Feb 19; 108(3):686-90 			

5.2 National Diabetes Audit (NDA)

Name and Description	Type of Data	Coverage	Geography	Duration
<p>The NDA is the largest annual clinical audit in the world, integrating data from both primary and secondary care sources, making it the most comprehensive audit of its kind</p> <p>It measures the effectiveness of diabetes healthcare against NICE clinical guidelines and NICE Quality standards in England and Wales</p>	Primary and secondary care clinical data	All ages; patients with a diagnosis of diabetes who attend speciality secondary care diabetes clinics or are registered at a participating GP practice before the audit end date (excludes gestational diabetes, impaired glucose tolerance and impaired fasting glucose)	England and Wales	Since early 2000s
Context of data collected	Volume	Events for data collected	Granularity	Coding
Data from primary care are routinely collected during patient consultations	In 2014-15 there were 1,894,887 people with diabetes in England and Wales; 57.3% of practices participated in the audit	<ul style="list-style-type: none"> • Registration and prevalence • Care processes • HbA1c results • Patient age distribution • Average age and BMI • Diabetes type distribution • Ketoacidosis episodes 	Per CCG, practice or specialist unit	Primary care Read codes
Extraction/Data collection	Consent	Access		
Data are collected from participating GP practices via an electronic data extraction tool through EMIS Web or SystemOne TPP, or manually from other electronic record systems	General practice – opt in Secondary care – all trusts with specialist diabetes services are expected to participate	<p>Open data files for 2010-11 can be downloaded from the website www.data.gov.uk</p> <p>More recent data: NDA data access requests should be submitted to the audit commissioner (HQIP – Health Quality improvement partnership) using the following link: http://www.hqip.org.uk/national-clinical-audit-and-patient-outcomes-programme-data-access-request-process/</p>		
Contact details	Example Publications			
diabetes@hscic.gov.uk	<ul style="list-style-type: none"> • Hill CJ et al., <i>Chronic kidney disease and diabetes in the National Health Service: a cross – sectional survey of the UK National Diabetes Audit</i> Diabet Med 2014 Apr;31(4):448-54 			

6 Health Surveys

6.1 General Practice Patient Survey

Name and Description	Type of Data	Coverage	Geography	Duration
<p>The GPPS covers aspects of patient experience of GP surgeries including experience of making an appointment, satisfaction with opening hours and the quality of care received from GPs and practice nurses</p> <p>Ipsos MORI administers the survey on behalf of NHS England. The questionnaires were developed by Ipsos MORI in conjunction with the University of Exeter Medical School and the General Practice and Primary Care Research Unit at the University of Cambridge</p> <p>The survey is sent out twice a year –in January and July</p>	Primary care survey	All adult patients registered with a GP in England may potentially be selected	England	From June 2007
Context of data collected	Volume	Events for data collected	Granularity	
<p>Patient details are obtained for each GP practice using registration records held on the HSCIC (Health and Social Care Information Centre) database</p> <p>A proportionately stratified, unclustered (random) sample is drawn from each practice</p>	For 2015, 854,032 surveys were completed (37.5% response rate)	<p>Access to GP services (e.g. by phone, online etc.)</p> <p>Seeing a preferred GP</p> <p>Making an appointment</p> <p>Waiting times</p> <p>Experience at last GP appointment</p> <p>Experience at last nurse appointment</p> <p>Opening hours</p> <p>Overall experience</p> <p>Managing own health</p> <p>State of own health today</p> <p>Care planning</p> <p>Out of hours services</p>	Surveys completed per patient Data available per GP practice	

Extraction/Data collection	Consent	Access
Surveys can be completed on paper, online, via a telephone helpline, in a language other than English, and in sign language	Implied consent by completion and return of the survey	www.gp-patient.co.uk/surveys-and-reports Individual practice level and CCG data can be freely downloaded from the website
Contact details	Linkage	Example Publications
GPPatientSurvey_Enquiry@ipsos.com	Could be linked with other primary care data by CCG/practice level codes	<ul style="list-style-type: none"> Ashworth M., et al <i>Antibiotic prescribing and patient satisfaction in primary care in England: cross-sectional analysis of national patient survey data and prescribing data</i>. Br J Gen Pract 2016 Jan;66(642):e40-6 Mounce LTA., et al <i>Establishing the validity of English GP Patient Survey items evaluating out-of-hours care</i> BMJ Qual Saf doi:10.1136/bmjqs-2015-004215

6.2 National Health Surveys

6.2.1 Health Survey for England

Name and Description	Type of Data	Coverage	Geography	Duration
Annual survey running since 1991 used by the government to plan health services and make important policy decisions – data on health, behaviour, social care, physical measures, mental health and wellbeing	Primary care survey	Adults aged 16 and over Children aged 2-15 from 1995	England	Started in 1991, annually
Context of data collected	Volume	Events for data collected	Granularity	
An interview with each eligible person in the household is followed by a nurse visit for those who agree to take part Addresses are chosen at random	Approx. 8000 adults and 2000 children per survey	<ul style="list-style-type: none"> • Core questions are asked each year on: • General health • Psychosocial indicators • Smoking, alcohol • Demographic and socioeconomic indicators • Use of health services • Measurements of height, weight, and BP • Each year there is also a focus on a population group, disease or condition 	Per individual participant	
Extraction/Data collection	Consent	Access		
The survey is carried out by NatCen Social Research	Participant opt in	Key findings available from www.hscic.gov.uk Data freely available http://www.hscic.gov.uk/catalogue/PUB19295 or link from https://www.ucl.ac.uk/hssrg/studies/hse		
Contact details	Example Publications			
The Data Archive University of Essex Wivenhoe Park Colchester Essex CO4 3SQ Tel 01206 872 001 Fax 01206 872 003	<ul style="list-style-type: none"> • Fraser S et al <i>Exploration of chronic kidney disease prevalence estimates using new measures of kidney function in the Health Survey for England</i> PloS One 2015 Feb 20; 10(2):e0118676 • Oyebode O et al., <i>Fruit and vegetable consumption and all-cause, cancer and CVD mortality: analysis of Health Survey for England data</i> J Epidemiol Community Health 2014 Sep;68(9):856-62 • Sims, M. et.al., <i>Did smoke free legislation in England reduce exposure to second-hand Smoke among non-smoking Adults? Cotinine Analysis from the Health Survey for England</i> Environ Health Perspect. 2012 Mar; 120(3): 425-430 • Jordan R, et al., <i>Passive smoking and chronic obstructive pulmonary disease: cross-sectional analysis</i> 			

6.2.2 Scottish Health Survey (SHeS)

Name and Description	Type of Data	Coverage	Geography	Duration
<p>The SHeS was introduced to provide a detailed picture of the health of the Scottish population in private households and is designed to make a major contribution to the monitoring of health in Scotland</p> <p>The SHeS aims to:</p> <ul style="list-style-type: none"> • estimate the occurrence of particular health conditions • estimate the prevalence of certain risk factors associated with health • look at differences between regions and between subgroups of the population • monitor trends in the population’s health over time • make a major contribution to monitoring progress towards health targets 	<p>Primary and social care</p>	<p>Adults (aged 16 and above) and children (0-15 years) living in private households in Scotland</p>	<p>Scotland</p>	<p>Since January 2008, annually</p>
Context of data collected	Volume	Events for data collected	Granularity	
<p>Participants are selected at random from an address list obtained from the post office</p> <p>Consenting participants are interviewed in their own home</p>	<p>In 2014, 4,659 adults and 1,668 children took part in the survey</p>	<p>Principle focus of the survey is cardiovascular disease and related risk factors</p> <p>General health and mental wellbeing</p> <p>Alcohol consumption</p> <p>Smoking</p> <p>Diet</p> <p>Physical activity</p> <p>Obesity</p> <p>Respiratory health and asthma</p> <p>Cardiovascular disease and diabetes</p> <p>Health risks and deprivation</p>	<p>Per participant</p>	

		Height and weight measurements Sub sample – blood pressure, waist circumference, urine and saliva
Extraction/Data collection	Consent	Access
1995-2011 – personal interview by trained interviewer, followed by nurse visit. Data collected by nurses included prescribed medication, anthropometric/biomedical measurements including lung function, blood sample 2012-2015 – personal interview by trained interviewer who asks questions using Computer Assisted Personal Interviewing (CAPI). Also collects blood pressure, saliva, urine samples and waist circumference.	Participant opt in	http://www.gov.scot/Topics/Statistics/Browse/Health/scottish-health-survey Data can be freely downloaded from the UK data service website Technical help is available for analysing the data via SPSS or SAS Datasets combining two or more years of data are available for large sample sizes
Contact details	Linkage	Example Publications
ScottishHealthSurvey@scotland.gsi.gov.uk Julie Landsberg 0131 244 2368 Craig Kellock 0131 244 2589 Daniel Adams 0131 244 5332 Analytical Services (HD) Population Health (ASD Health) Basement Rear, St. Andrews House Regent Road Edinburgh EH1 3DG	Can be linked to Scottish morbidity records e.g. cancer registry data	<ul style="list-style-type: none"> • Kyle RG et al., <i>Prevalence of overweight and obesity among nurses in Scotland: A cross – sectional study using the Scottish Health Survey</i> Int J Nurs Stud 2016 Jan;53:126-33 • Watt RG., <i>Tooth loss and cardiovascular disease mortality risk – results from the Scottish Health Survey</i> PLoS One 2012;7(2):e30797 • Leung J <i>Psychological distress, optimism and general health in breast cancer survivors: a data linkage study using the Scottish Health Survey</i> Support Care Cancer 2016 Apr; 24(4): 1775-61

6.2.3 Welsh Health Survey (WHS)

Name and Description	Type of Data	Coverage	Geography	Duration
<p>The survey provides information on health status, illnesses, lifestyle, health survey use and children</p> <ul style="list-style-type: none"> • The information collected is used in many ways including to • Provide national estimates of health and health-related lifestyle • Examine differences between population sub group and local areas • Provide evidence and monitor targets for promoting better health • Provide information for development of health, social care and wellbeing strategies <p>After 2015 the Welsh Health Survey will be incorporated in the National Survey for Wales which will include health related questions</p>	Self-reported primary and social care	Adults and children living in private households in Wales, to include a minimum of 600 adults from each local authority area	Wales	From 2003 – 2015 published annually
Context of data collected	Volume	Events for data collected	Granularity	Coding
<p>Participants are selected at random using the Post Office Postcode Address File</p> <p>The sample is stratified by local authority</p> <p>The survey collects information on households (through a short interview) and individuals (through a self-completion questionnaire)</p> <p>At each household all adults and a maximum of two children are eligible for inclusion in the survey</p>	Approx. 15,000 adults and 3,000 children per year	<p>Adults:</p> <ul style="list-style-type: none"> • General health and wellbeing • Health service use • Medicines and illnesses • Untreated problems or symptoms • Alcohol and smoking • Fruit and vegetable consumption • Exercises • Carers • Height and weight • Demographics 	Individuals Families/households	Self-completion questionnaires are edited using NatCen's in-house system Data are numeric

		<ul style="list-style-type: none"> • Infant feeding <p>For children:</p> <ul style="list-style-type: none"> • Health status • Health service use • Accidents • Illnesses and other health problems • Eating habits • Physical activity • Strengths and difficulties (SDQ) 		
Extraction/Data collection	Consent	Access		
Interviewers conduct a short interview with a responsible adult in the household, covering basic socio-demographic information Self-completion questionnaires are then left for completion by all adults in the household for later collection	Participant opt in	http://gov.wales/statistics-and-research/welsh-health-survey/ https://statswales.wales.gov.uk/Catalogue/Health-and-Social-Care/Welsh-Health-Survey Data can be downloaded from the UK data service www.ukdataservice.ac.uk		
Contact details	Example Publications			
stats.healthinfo@wales.gsi.gov.uk	<ul style="list-style-type: none"> • Shiue I and Sand M., <i>Quality of life in caregivers with and without chronic disease: Welsh Health Survey, 2013</i>, J Public Health (Oxf.) 2016 Jan 25 			

6.2.4 Health Survey Northern Ireland

Name and Description	Type of Data	Coverage	Geography	Duration
<p>Health Survey Northern Ireland is a Department of Health, Social Services and Patient Safety that runs every year on a continuous basis</p> <p>The survey covers a range of health topics Separate modules for different policy areas are included in different financial years</p>	Survey – health and social care	Children and adults aged 16 and over living in private households in Northern Ireland.	Northern Ireland	Since 2010
Context of data collected	Volume	Events for data collected	Granularity	
A random sample of 5850 addresses across Northern Ireland were selected for interviewing in 2014/5	2014/15 survey based on 4,144 individual responses	<ul style="list-style-type: none"> • General health • Mental health and wellbeing • Diet and nutrition • Breastfeeding • Oral health • Medicines • Obesity (Body mass index) • Smoking • Sexual health 	Per household or participant	
Extraction/Data collection	Consent	Access		
Data were collected using Computer Assisted Personal Interviewing (CAPI) and Computer Assisted Self Interviewing (CASI)	Participant opt in	https://www.health-ni.gov.uk/topics/doh-statistics-and-research/health-survey-northern-ireland Questionnaires and excel data files available for download from website 2010-11 Data available from UK Data service		
Contact details	Example Publications			
phirb@health-ni.gov.uk	<ul style="list-style-type: none"> • Dee A., et al <i>Overweight and obesity on the island of Ireland: an estimation of costs</i> BMJ Open 2015; 5:e006189 doi:10.1136 • Bhatnagar P. et al., <i>The epidemiology of cardiovascular disease in the UK 2014</i> Heart doi:10.1136/heartjnl-2015-307515 			

6.3 Cancer Patient Experience Surveys

Name and Description	Type of Data	Coverage	Geography	Duration
<p>England: National Cancer Patient experience survey Wales: Welsh Cancer Patient Experience N. Ireland: Northern Ireland Cancer Patient Experience Survey</p> <p>The Cancer Reform Strategy (CRS) published in 2007 set out a commitment to establish a new NHS Cancer Experience Survey Programme to monitor national progress and to use the findings to drive quality improvements locally</p>	Patient experience survey	Patients seen for cancer treatment in hospital	England Wales N. Ireland	England: since 2010 Wales: since 2013 N. Ireland: since 2014
Context of data collected	Volume	Events for data collected	Granularity	
The surveys cover acute and specialist NHS Trusts that provide adult acute cancer services	England: >70,000 patients responded to the survey in 2014 Wales: sent to 11,000 patients in 2013 N. Ireland: >5000 in 2014-5	<ul style="list-style-type: none"> • Results by tumour groups • Patients' experiences re; • Seeing GP • Diagnostic tests • Finding out diagnosis • Deciding best treatment • Clinical nurse specialists • Support for patients • Operations • Hospital Doctors • Ward nurses • Hospital care and treatment • Information before leaving and home support • Day/outpatient care • Outpatients appointments • Care from general practices • Overall NHS care • First treatment and occurrence 	Per patient	

Extraction/Data collection	Consent	Access
<p><u>England</u>: in 2014, the survey covered patients seen for treatment in hospital between 1st September 2013 and 30th November 2013, who had a primary diagnosis of cancer</p> <p><u>Wales</u>: the survey covered 11,000 adult cancer patients who had been admitted to hospital, seen as day case or discharged from 1st June 2012 to 31st March 2013</p> <p><u>N. Ireland</u>: the survey covered approximately 5000 adult cancer patients who had been inpatients or day case patient and who were in active treatment between 1st December 2013 and 31st May 2014</p>	<p>Implied consent on return of survey</p>	<p>www.quality-health.co.uk/surveys</p> <p>Data and survey results for English and M. Ireland survey are freely available and can be downloaded from the website</p> <p>Experience reports from Wales in Welsh language on website</p>
Contact details	Linkage	Example Publications
<p>Surveys are carried out by Quality Health</p> <p>Address: Quality Health Limited, Unit 1, Holmewood Business Park, Chesterfield Road, Holmewood, Chesterfield, Derbyshire S42 5US, United Kingdom</p> <p>Telephone: 01246 856263</p> <p>Fax: 01246 855897</p> <p>Email: info@quality-health.co.uk</p>	<p>Surveys from each nation can be compared with each other</p>	<ul style="list-style-type: none"> • Saunders CL, et al., <i>What explains worse patient experience in London? Evidence from secondary analysis of the Cancer Patient Experience Survey</i>. BMJ Open 2014;4:e004039 doi:10.1136/bmjopen-2013-004039 • Saunders CL, et al., <i>Accuracy of routinely recorded ethnic group information compared with self-reported ethnicity: evidence from the English Cancer Patient Experience survey</i>. BMJ Open. 2013;3(6). pii: e002882. doi: 10.1136/bmjopen-2013-002882 • Lyratzopoulos G, et al., <i>Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England</i>. Lancet Oncol. 2012;13(4):353-65

6.4 Scottish Diabetes Surveys

Name and Description	Type of Data	Coverage	Geography	Duration
The Scottish Diabetes survey describes many aspects of diabetes care across the whole of Scotland The survey runs annually	Primary and secondary care clinical data	General practices and hospitals in Scotland	Scotland	Since 2002
Context of data collected	Volume	Events for data collected	Granularity	
Data are collected during routine primary and secondary care consultations	SCI-Diabetes holds data from 994 General practices, 42 hospitals and 13 regions linking to local laboratory data (approx. 276,00-patients)	Demographics Incidence and prevalence of diabetes Duration of diabetes Mortality Type of diabetes Glycaemic control Blood pressure, cholesterol, smoking status Complications of diabetes – cardiovascular disease, kidney disease, eye disease, foot complications	Per registered patient	
Extraction/Data collection	Consent	Access		
Most of the data contained in the survey is extracted from the diabetes IT system SCI (Scottish Care Information)-Diabetes Data on SCI-Diabetes can be viewed by Scottish NHS GP practices, hospital teams and patients in Scotland through www.mydiabetesmyway.scot.nhs.uk	Not required (data collection similar to audit)	Survey data can be downloaded from: http://www.diabetesinscotland.org.uk/Publications.aspx?catId=3		
Contact details	Linkage	Example Publications		
Contact for website Boris Rogatchevski Boris.rogatchevski@gov.scot	Can be linked to other data via the Community Health Index (CHI) number	<ul style="list-style-type: none"> McKnight JA, et al., <i>Implementing a national quality assurance system for diabetes care: the Scottish Diabetes Survey 2001-2006</i> Diabet Med 2008 June;25(6):743-6 		

7 Special Datasets

7.1 Aberdeen Maternity and Neonatal Databank

Name and Description	Type of Data	Coverage	Geography	Duration
The Aberdeen Maternity and Neonatal Databank was initiated in the department of obstetrics and gynaecology by the late Professor Sir Dugald Baird, In 1950 From 1951 to present this database links all the obstetric and fertility related events occurring in women from a defined population	Clinical data collected in primary and secondary care	Aberdeen city pregnancy and births	Aberdeen, Scotland	Since 1950
Context of data collected	Volume	Events for data collected	Granularity	
Data are collected from maternity medical notes	Total number of women: 181,145 Pregnancies: 274,323 Deliveries: 272,449 Babies: 277,324 Twins: 4,120	<ul style="list-style-type: none"> • Patient details • Pregnancy details • Delivery details • Baby details • Other variables including diagnostic procedures, operative management, drugs, indications for procedures, intergenerational data 	Per patient	
Extraction/Data collection	Consent	Access		
Updated on a weekly basis from a separate continuously maintained database, which holds medical data as free text at the point of data entry. The database is updated by up to 5 concurrent users with information relating to approximately 100 births every week	A routine opt-out clause is applicable at time of antenatal booking	All requests for access must be made through the Databank Steering Committee A databank request application form can be downloaded from the website http://www.abdn.ac.uk/iahs/research/obsgynae/amnd/access.php		

Contact details	Linkage	Example Publications
<p>Dr. Sohinee Bhattacharya Dugald Baird Centre for Research on Women's Health Aberdeen Maternity Hospital Cornhill Road, Aberdeen AB25 2ZL Tel: +44 (0)1224 438441 Fax: +44 (0)1224 438486 e-mail: sohinee.bhattacharya@abdn.ac.uk</p>	<p>Can be linked to other Scottish records using CHI numbers</p>	<ul style="list-style-type: none"> • Taylor CM et al., <i>Folic acid in pregnancy and mortality from cancer and cardiovascular disease: a further follow-up of the Aberdeen folic acid supplementation trial</i> J Epidemiol Community Health 2015, 0, 1-6 • Lee KK, et al., Maternal obesity during pregnancy associates with premature mortality and major cardiovascular events in later life. Hypertension 2015 Nov; 66(5):938-44

7.2 Scottish Health Informatics Centre (HIC) Services

Name and Description	Type of Data	Coverage	Geography	Duration
<p>HIC Services is a University of Dundee research support unit within the Tayside medical Science Centre (TASC) and the Farr Institute @ Dundee in collaboration with NHS Tayside and NHS Fife</p> <p>HIC Services operates a secure Safe Haven environment with strong data governance for the provisioning of data to academics and other users to improve healthcare and population health</p> <p>A large number of datasets exist which can be viewed via the website including:</p> <ul style="list-style-type: none"> Acute hospital admissions Prescribing data Laboratory data Death registrations Cancer registrations Disease registers e.g. Diabetes diagnoses Research datasets e.g. Scottish Dementia Clinical Research Network (SDCRN) dataset and Scottish Heart Health Study 	Majority are secondary care datasets	In general, the population of Tayside and Fife who have a community health index (CHI) number	Scotland	Datasets vary in duration, with some extending over 15 years+
Context of data collected	Volume	Events for data collected	Granularity	
The majority of data are routinely collected	Data covers approx. 800,000 people, about 16% of Scottish population	Varies by dataset HIC hosts and updates linkable health data for: Community-dispensed prescribing Hospital stays, diagnoses and interventions Person based laboratory data – e.g. biochemistry, haematology, microbiology, virology and immunology Registrations e.g. deaths, cancers	Generally, per registered patient	

Extraction/Data collection	Consent	Access	Coding
<p>HIC maintains up-to-date prescribing and hospitalisation data through regular updates and data feeds from suppliers</p>	<p>All medical data is anonymised and consent is not necessary; data is managed safely in compliance with Data protection legislation</p>	<p>http://medicine.dundee.ac.uk/hic See document – HIC Services: User Guide – link on website Costs apply depending on what data are required Researchers should initially contact Jim Galloway for a scoping meeting Data are released for use in the Safe Haven – data are not released externally but placed on a server at HIC, within a restricted, secure IT environment NHS Tayside and NHS Fife R&D approval must be applied for when using health board data</p>	<p>Will vary according to dataset Prescribing: BNF; Medica Conditions: ICD10 and ICD09; Procedures: OPCS v4; Lab data is locally coded</p>
Contact details	Linkage	Example Publications	
<p>HIC Services The Farr Institute University of Dundee (Main level 5 corridor), Second Floor, Level 7 Mailbox 15 Ninewells Hospital & Medical School Dundee, DD19SY hicservices@dundee.ac.uk Data linkage: hicsupport@dundee.ac.uk</p>	<p>Data Linkage Service (DLS) is available via NHS or CHI (Community Health Index) numbers HIC can link to NHS or non NHS datasets not already hosted by HIC</p>	<ul style="list-style-type: none"> • Duffy MA et al, <i>Trimethoprim prescription and subsequent resistance in childhood urinary infection: multilevel modelling analysis</i> Br J Gen Pract 2013 Apr;63(609):e238-43 • Lockhart P and Guthrie B <i>Trends in primary care antidepressant prescribing 1995-2007: a longitudinal population database analysis</i> Br J Gen Pract 2011 Sep;61(590):e565-72 • Guthrie B et al., <i>High risk prescribing in primary care patients particularly vulnerable to adverse drug events: cross sectional population database analysis in Scottish general practice</i> BMJ 2011 Jun 21;342:d3514 	

8 Cohort studies

8.1 National

8.1.1 UK Biobank

Name and Description	Type of Data	Coverage	Geography	Duration
<p>Cohort study</p> <p>Major national health resource with aim of improving prevention, diagnosis and treatment of serious and life-threatening illness</p> <p>Hosted by the University of Manchester</p>	<p>Primary and secondary care data</p>	<p>Participants who consented to take part in the study</p>	<p>UK wide</p>	<p>Participants were recruited between 2006 and 2010</p>
Context of data collected	Volume	Events for data collected	Granularity	Coding
<p>Participants agreed to have data collected about them</p>	<p>500,000 participants between 40 and 60 years</p>	<ul style="list-style-type: none"> • Data collected at assessment visit included questionnaires on health, lifestyle, diet, memory, work, family history • Physical measures: blood pressure, arterial stiffness, eye measurements (visual acuity, intraocular pressure, optical coherence tomography), body composition measurements, hand grip strength, ultrasound bone densitometry, spirometry, exercise ECG, resting ECG, physical activity monitor, • Sample of blood, urine and saliva were collected • Web based dietary questionnaire was included in assessment visit towards the end of the recruitment periods • Genetic data • Some original participants have been invited back for brain, heart and body imaging e.g. brain MRI, liver MRI, cardiac MRI, carotid 	<p>Per registered participant</p> <p>Summary data available on website</p>	<p>Uses Data-field codes (a data-field is fundamental block of data held within the Biobank repository and identifies the results of a single question, measurement or result)</p>

		ultrasound		
Extraction/Data collection	Consent	Access		
Participants were recruited through assessment centres and data collected at assessment Some participants have been invited back for further testing	Participant opt in Can withdraw consent at any time	www.ukbiobank.ac.uk Researchers can apply to use the data for health-related research that is in the public interest Application process in two stages – preliminary and main To register and apply researchers should access the link on the website to the access management system (AMS) No limit on data or samples which can be requested but must be relevant to application Quote for data will be given when preliminary information received Separate ethical approval may not be required as UK Biobank has approval which covers majority of uses		
Contact details	Linkage	Example Publications		
0800 0276 276 access@ukbiobank.ac.uk	UK Biobank provides de-identified data but recognises that in certain circumstances two or more researcher groups may want to link their datasets together. UK Biobank can help with proposed linkage meeting certain criteria	<ul style="list-style-type: none"> • Walker-Bone K. et al., <i>Chronic widespread bodily pain is increased among individuals with history of fracture: findings from UK Biobank</i>, Arch Osteoporosis 2016; 11:1 • Tyrell J. et al., <i>Height, body mass index and socioeconomic status: Mendelian randomisation study in UK Biobank</i>, BMJ 2016;352:i582 • Flint E and Cummins S., <i>Does active commuting protect against obesity in mid-life? Cross-sectional, observational evidence from UK Biobank</i>, The Lancet 2015, 386, S8 • Guggenheim J.A., et al., <i>Childhood febrile illness and the risk of myopia in UK Biobank participants</i> Eye(Lond) 2016 Feb 5 (Epub ahead of print) 		

8.1.2 MRC National Survey for Health and Development (NSHD)

Name and Description	Type of Data	Coverage	Geography	Duration
<p>The MRC National Survey of Health and Development is the oldest of the British Cohort studies and has data from birth on the health and social circumstances of a representative sample of men and women born in March 1946</p> <p>The MRC Unit for Lifelong Health and Ageing at UCL (LHA) is responsible for the study</p>	Cohort study	Men and women born during one week in England, Scotland or Wales in March 1946	England, Scotland, Wales	Since 1946
Context of data collected	Volume	Events for data collected	Granularity	Coding
<p>Data on study members collected via contacts with the study members, questions answered by parents, school-teachers, school doctors and nurses, and youth employment officers. From aged 0-15 information collected from parent. Between aged 5-15 some information was collected from study members themselves, usually in form of cognitive tests. From aged 16, members were interviewed personally.</p> <p>Information has also been collected from other sources e.g.</p> <ul style="list-style-type: none"> • Reports of hospital admissions • Information on exposure to outside atmospheric pollution from birth to age 11 • Details of educational qualifications • Notifications of deaths • Cancer diagnosis • Mammograms 	5362 participants	<p>There are approximately 24,000 variables in the repository including;</p> <ul style="list-style-type: none"> • Physical growth and development; changes in adult body size and physical function • Cognitive development and cognitive ageing • Lifetime physical and mental health and health related behaviours • Indicators of physical and social environment • Cardiac and vascular imaging, whole body DXA scans and pQCT scans of the radius <p>Biological samples:</p> <ul style="list-style-type: none"> • Buccal and blood samples at 53 years • Blood, urine, and saliva samples at 60-64 years • Blood samples at 69 years • (DNA extracted various biomarkers including 'omics) 	Per registered participant	ICD and BNF coding systems used to code applicable variables e.g. medications and medical conditions

Extraction/Data collection	Consent	Access	
There have been 24 waves of data collected on the whole cohort between 1946 and 2015 as well as smaller collections from various subgroups	Parents gave consent at time of birth; participant opt in for adult studies	<p>www.nshd.mrc.ac.uk</p> <p>Access to data are permitted via data sharing agreement between the applicants institution and the LHA</p> <p>The meta-data repository holds electronic versions of all NSHD questionnaires and topic level guides to information collected.</p> <p>Interested researchers can use the repository to explore the NSHD meta-data to help formulate research proposals</p> <p>Interested researchers should email the contact address for further information and informal advice as to whether the research question can be answered using NSHD data</p> <p>There are currently no costs to researchers associated with supplying or accessing NSHD data</p>	
Contact details	Linkage	Example Publications	Ongoing studies
mrclha.swiftinfo@ucl.ac.uk	NSHD data can be linked to ONS mortality and cancer data and HES data	<ul style="list-style-type: none"> • Cooper A. et al., <i>Physical Activity, Sedentary Time and Physical Capability in Early Old Age: British Birth Cohort Study</i> PLoS One 2015; 10(5): e0126465 • Richards M. et al <i>Lifetime affect and midlife cognitive function: prospective birth cohort study</i> Br J Psychiatry 2014 Mar; 204(3): 194-199 • Mishra G and Kuh D. <i>How do health symptoms during midlife relate to menopausal transition? A British prospective cohort study</i> BMJ 2012; 344:e402 • Kuh D, et al <i>Cohort Profile: updating the cohort profile for the MRC National Survey for Health and Development: a new clinic-based data collection for ageing research</i> Int J Epidemiology 2011 February 40(1): e1-e9 	<p>On-going studies include:</p> <ul style="list-style-type: none"> • The lifetime determinants of cardiovascular, musculoskeletal and mental ageing and wellbeing • Puberty, menopause and ageing phenotypes • Social and psychological studies of resilience • Genetic, epigenetic and metabolomics studies

8.1.3 Boyd Orr Cohort

Name and Description	Type of Data	Coverage	Geography	Duration
The Boyd Orr cohort is based on a cross-sectional survey of childhood diet and health involving children in the 1930s. It is an historical cohort study carried out by the University of Bristol Department of Social Medicine to investigate the long term impact of children's diet, growth, living conditions and health based on adult cardiovascular disease.	Cohort study (childhood diet and health)	4999 children (aged 0-16) examined in 16 centres between 1937 and 1939	England and Scotland	Original survey took place in 1930s
Context of data collected	Volume	Events for data collected	Granularity	
The study is based on the long term follow up of children who took part in the Carnegie Survey of Diet and Health in Pre-War Britain (1937-9)	4999 children	<ul style="list-style-type: none"> Family diet and health in pre-war Britain Follow up diet and health 1997-8 Follow up diet and health 2002-3 Arterial ultrasound Blood samples 	Per registered participant	
Extraction/Data collection	Consent	Access		
Initial survey of 4999 children. Further surveys were sent to 3220 participants in 1997/98. In 2002/3 a further survey was sent and over 700 participants gave blood samples (including for DNA extraction) and several hundred had arterial ultrasound scans	Participant consent for 1997 questionnaire survey and 2002 research clinics	www.bris.ac.uk/social-community-medicine/projects/boyd-orr/ Researchers interested in collaborating should contact the current team (see contacts) and complete the collaborator's agreement.		
Contact details	Linkage	Example Publications		
Boyd Orr Cohort Study School of Social and Community Medicine University of Bristol Canyng Hall 39 Whatley Road Bristol, BS8 2PS Professor Richard Martin Richard.martin@bristol.ac.uk Professor George Davey Smith kz.davey-smith@bristol.ac.uk Professor David Gunnell d.j.gunnell@bristol.ac.uk	This cohort study forms part of the HALCyon combined cohort study	<ul style="list-style-type: none"> Martin RM. et al <i>Cohort profile: The Boyd Orr cohort – an historical cohort study based on the 65 year follow-up of the Carnegie Survey of Diet and Health (1937-39)</i> Int J Epidemiol 2005 Aug; 34(4): 742-9 Martin RM., et al., <i>Breastfeeding and cancer: The Boyd Orr cohort and a systematic review with meta-analysis.</i> J Natl Cancer Inst 2005 97(19): 1446-1457 Montgomery S. et al., <i>Pre-pubertal growth and blood pressure in early old age</i> Archives of Childhood 2000; 82:358-363 		

8.1.4 HALCyon (Healthy Ageing across the Life Course) cross-cohort collaboration

Name and Description	Type of Data	Coverage	Geography	Duration
<p>HALCyon brought together 9 UK cohort studies for comparative research on factors across life associated with markers of healthy ageing, namely physical and cognitive capability and psychological wellbeing.</p> <p>These cohort studies have followed up groups of people to gather repeated measures of life experience and status. Together they include 30,000 men and women born between 1918 and 1958 who were <i>50 years and older</i> at the start of the programme in 2008.</p> <p>The studies are:</p> <ul style="list-style-type: none"> • Lothian Birth Cohort 1921 • Hertfordshire Cohort Study • Hertfordshire Ageing Study • Boyd Orr Cohort study • Aberdeen Birth Cohort 1936 • MRC National Survey of Health and Development (1946 British Birth Cohort) • National Child Development Study (1958 British Birth Cohort) • English Longitudinal study of Ageing (ELSA) • Caerphilly Prospective Study 	<p>Combined cohort studies</p>	<p>Individuals born between 1918 and 1958 who were 50 years or older at the start of the programme in 2008</p>	<p>England, Scotland and Wales</p>	<p>HALCyon commenced in 2008; individual cohort studies from 1921 onwards</p>
<p>Context of data collected</p>	<p>Volume</p>	<p>Events for data collected</p>	<p>Granularity</p>	

The collaboration combines data from the 9 cohort studies. Seven are life course studies with data from childhood and adulthood. ELSA and Caerphilly are adult cohorts.	30,000 individuals born between 1918 and 1958	All cohorts have data on adult physical or cognitive capability All have measures of self-reported functioning, wellbeing and mental health, lifetime social conditions, adult lifestyle. Most also have DNA, tests and serum markers of cardiovascular and biological function, stored blood samples. Five have measures of infant and/or childhood growth Four have measures of childhood cognitive function Two have data on childhood diet	Per participant	
Extraction/Data collection	Consent	Access	Coding	
Data collected as per individual cohort studies Secondary data from the 9 participating studies was harmonised	As per individual cohort study	http://www.nshd.mrc.ac.uk/collaborations/halcyon/ Lists of the comparable data are freely available on the HALCyon data documentation page (via website) Potential researchers can apply to access data on wellbeing, telomere length and cortisol by completing the HALCyon data sharing form (via website)	Stata syntax was used to clean and recode data for use in analyses, comparable across cohorts	
Contact details	Linkage	Example Publications		
Stephanie.Pilling@ucl.ac.uk	The 9 cohort studies are not linked	<ul style="list-style-type: none"> Cooper R. et al., <i>Physical capability and subsequent positive mental wellbeing in older people: findings from five HALCyon cohorts</i> Age 2014; 36:445-456 Gale CR., et al <i>Factors associated with symptoms of anxiety and depression in five cohorts of community-based older people: the HALCyon (Healthy ageing across the life course) programme</i> Psychological Medicine 2011; 410 (10): 2057 - 73 		

8.2 Regional

8.2.1 Born in Bradford (England)

Name and Description	Type of Data	Coverage	Geography	Duration
<p>Cohort study led by the Universities of Bradford, Leeds and York as well as other universities and stakeholders</p> <p>By recruiting pregnant women, their partners and new-born babies this study offers the potential to:</p> <ul style="list-style-type: none"> Assess the determinants of childhood and adult disease Assess the impact of migration Explore the influences of pregnancy and childbirth on subsequent health Generate and test hypotheses that have the potential to improve health for some of the most disadvantaged in society 	Primary care and secondary care	Participants who agreed to take part in the study	Bradford, England	<p>Recruitment between 2007 and 2011</p> <p>Data collection on-going</p>
Context of data collected	Volume	Events for data collected	Granularity	
Long term cohort study of 13,500 children born at Bradford Royal Infirmary between March 2007 and December 2010	Details from 12,543 mothers with 13,776 pregnancies and 3448 of their partners	<ul style="list-style-type: none"> Mother's weighed and measured Infants – details anthropometric assessment at birth to 2 years Oral glucose tolerance test and lipid profiles obtained at 28 weeks Pregnancy serum, plasma and urine samples stored Cord blood samples stored and DNA on 10000 mother-infant pairs Other events available on website e.g. household composition, education, nutrition social factors Ongoing sample collection as per nested study requirements 	Per participant	

Extraction/Data collection	Consent	Access
<p>Health of participants tracked from pregnancy through childhood and into adult life</p> <p>The study relies on the collection of routine measurements by paediatricians, midwives, school nurses, GPs and health visitors</p> <p>Ongoing data collection via biological samples, height, weight, and skinfold measurements for platform studies and nested studies</p>	Participant opt in	<p>www.borninbradford.nhs.uk</p> <p>Researchers need to read 'Guidance for Collaborators' and complete expression of interest form available on website</p> <p>Proposals are reviewed on a monthly basis</p>
Contact details	Linkage	Example Publications
<p>Rosie.mceachan@bthft.nhs.uk</p>	<p>Born in Bradford data can be linked to primary care GP data via TPP ResearchOne.</p>	<ul style="list-style-type: none"> • Uphoff E. et al., <i>A systematic review of socioeconomic position in relation to asthma and allergic diseases</i>, European Respiratory Journal 2015 47;364-374 • Bryant, M. et al., <i>Agreement between routine and research measurement of infant height and weight</i>, Archives of Disease in Childhood, 2015 100(1): 24-29 • West J. et al., <i>Differences in socioeconomic position, lifestyle and health-related pregnancy characteristics between Pakistani and White British women in the Born in Bradford prospective cohort study: the influence of the woman's, her partner's and their parents' choice of birth</i>, BMJ Open 2014 4(6):e004805

8.2.2 Aberdeen Birth Cohorts (Scotland)

Name and Description	Type of Data	Coverage	Geography	Duration
<p>The University of Aberdeen has followed all of the children born in Aberdeen in 1921, 1936, and 1950-56 as they grow up</p> <p>There are 3 Cohorts; 1921 Birth Cohort 1936 Birth Cohort Children of the 1950s</p>	<p>Primary health and social care</p>	<p>Children born in Aberdeen in 1921, 1936 and between 1950-56</p>	<p>Aberdeen, Scotland</p>	<p>Since 1921</p>
Context of data collected	Volume	Events for data collected	Granularity	
<p>In primary school the children sat tests of reading and mental ability</p>	<p>1950s - 12,150 children born in Aberdeen between 1950-56 1936 – 350 children born in 1936 1921 – 87,498 school children born in 1921</p>	<ul style="list-style-type: none"> • Reading and mental ability tests • Questionnaires on current health and social topics • Blood tests and brain imaging for some participants 	<p>Per registered participant</p>	
Extraction/Data collection	Consent	Access		
<p>In the 1990s participants of the 1950s cohort were sent a postal questionnaire giving information on current health and other topics Participants of the 1921 and 1936 cohorts have also been followed up Over the years, researchers have linked the results from these tests to health and social information including blood samples and brain imaging</p>	<p>Patient/parent consent</p>	<p>http://www.abdn.ac.uk/birth-cohorts/1950s/for-researchers/data-access Access to the data are given by application to the steering committee Participants' information is shared anonymously with researchers and used in a data safe haven Draft application form available on the website</p>		

Contact details	Linkage	Example Publications
Heather Clark h.clark@abdn.ac.uk +44 (0)1224 438443	Test results were linked to other school records and to birth records in the Aberdeen Maternity and Neonatal Databank.	<ul style="list-style-type: none"> • Dundas R et al., <i>Early life school, neighbourhood and family influences on adult health: a multilevel, cross-classified analysis of the Aberdeen Children of the 1950s study</i>. Am J Epidemiol 2014 Jul 15; 180(2):197-207 • Lawlor DA et al., <i>Obstetrician-assessed maternal health at pregnancy predicts offspring future health</i>. PLoS ONE. 2007 Aug 1;2:e666 • Lawlor DA, et al., <i>The association of birth weight, gestational age and childhood BMI with type 2 diabetes: findings from the Aberdeen Children of the 1950s cohort</i> Diabetologia. 2006 Nov;49(11):2614-7. Epub 2006 Sep 21

8.2.3 Growing up in Scotland

Name and Description	Type of Data	Coverage	Geography	Duration
<p>Growing up in Scotland is a longitudinal research study tracking the lives of thousands of children and their families from the early years, through childhood and beyond</p> <p>The longitudinal nature of the data is particularly useful in looking at how early experiences influence later outcomes</p>	<p>Cohort study (health, education, social care)</p>	<p>Nationally representative sample of families in Scotland with children of particular ages.</p>	<p>Scotland</p>	<p>From 2005, data collection ongoing</p>
Context of data collected	Volume	Events for data collected	Granularity	
<p>Three cohorts of children have been taking part in GUS: Child Cohort - approx. 3000 children born between June 2002 and May 2003</p> <p>Birth Cohort 1 – approx. 5000 children born between June 2004 and May 2005</p> <p>Birth Cohort 2 – approx. 6000 children born between March 2010 and February 2011</p> <p>Families were selected at random from Child Benefit records provided by HMRC.</p>	<p>Approx. 14,000 children across three cohorts</p>	<ul style="list-style-type: none"> • Cognitive, social, emotional and behavioural development • Physical and mental health and wellbeing • Childcare, education and employment • Home, family, community and social networks • Involvement in offending and risky behaviour 	<p>Per child</p>	
Extraction/Data collection	Consent	Access	Coding	
<p>Data is collected from families at different time points</p> <p>Data is collected through in-home visits and encompasses both interview data and objective measurements (incl. child’s cognitive development and height and weight measurements). The main interview data is collected from the child’s main carer using Computer Assisted Personal Interviewing (CAPI) and Computer Assisted Self Interviewing (CASI). From age 8 onwards interview data is also collected from the cohort child.</p>	<p>Participation is voluntary</p>	<p>www.growingupinScotland.org.uk</p> <p>GUS data can be accessed free of charge from the UK Data service (link from website)</p> <p>It is available in SPSS or STATA</p> <p>Workshops on data access available</p>	<p>Unique – user guide and overview of data available</p>	

Contact details	Linkage	Example Publications	
<p>Line Knudsen, Senior Researcher, ScotCen Social Research Line.Knudsen@scotcen.org.uk</p> <p>Paul Bradshaw, Head of ScotCen Social Research Paul.bradshaw@scotcen.org.uk</p>	<p>Survey data has been linked to child's education records and to child and maternal health records (contact the research team for details)</p>	<p>Parkes A., et al., <i>Does parenting help to explain socioeconomic inequalities in children's body mass index trajectories? Longitudinal analysis using the Growing up in Scotland Study</i> J Epidemiol Community Health 2016 Apr 7</p>	

9 Administrative datasets

9.1 Scottish Administrative Records

Name and Description	Type of Data	Coverage	Geography	Duration
<p>The information Services Division (ISD) is part of NHS Scotland and collects, manages and centrally holds a variety of Scottish Health Data</p> <p>Main datasets are:</p> <ol style="list-style-type: none"> 1) Outpatient Attendance Dataset (SMR00) (from 1997) 2) General Acute/Inpatient dataset (SMR01) (from 1981) 3) Maternity inpatient and day cases dataset (SMR02) from 1981 4) Mental Health Inpatient and Day case dataset (SMR04) from 1981 5) Scottish Cancer Registry (SMR06) 6) Neonatal inpatient dataset (SMR11) 7) Scottish Drug Misuse Database 8) Scottish Birth Record 9) Maternity and Neonatal Linked Database 10) Scottish Morbidity Database 11) Prescribing information system 	Administrative data	Scotland	Scotland	Varies by dataset but earliest from 1981
Context of data collected	Volume	Events for data collected	Coding	
<p>Administrative data are collected by ISD on behalf of the NHS Scotland</p> <p>For example, the Outpatients (SMR00) dataset collects episode level data from patients on new and follow up appointments at outpatients clinics in all specialities.</p>	<p>Varies by dataset</p> <p>SMR00 – approx. 4.4.million records per year</p> <p>SMR01 – approx. 1.4 million records per year</p> <p>SMR02 – approx.120, 000 records per year</p> <p>SMR04 – approx. 21,000 records per year</p> <p>SMR06 – 1,400,000</p>	<p>Varies by dataset including:</p> <ul style="list-style-type: none"> • demographic data • admission/appointment type • waiting times • patient’s diagnoses, conditions • procedures/operations • pregnancy and birth details 	<p>Varies by dataset; includes Read, ICD-10, OPCS-R and SNOMED-CT</p>	

	<p>records (approx. 47,000 records added each year)</p> <p>SMR11 – approx. 1,300,000 records</p> <p>Scottish Drug misuse – approx. 12,000 new and 13,000 follow up assessments per year</p> <p>Scottish birth record</p> <p>Maternity and Neonatal linked database – approx. 60,000 records per year</p> <p>Scottish Morbidity database – over 37 million records</p> <p>PIS- around 90 million data items per annum</p>	<ul style="list-style-type: none"> • cancer site and stage 	
Extraction/Data collection	Consent	Access	
<p>Data are collected routinely according to each dataset</p> <p>Health boards supply data to ISD on a continual basis throughout the year and have targets to submit their data within six weeks</p> <p>The dataset is generally fully complete and ready for analysis three months preceding the current date</p> <p>The neonatal database is updated every six months</p>	<p>Not possible, but complies with Data Protection Act</p>	<p>http://www.adls.ac.uk/find-administrative-data/introduction-to-nhs-scotland/</p> <p>ISD has developed the electronic Data Research and Innovation Service (eDRIS) to assist researchers:</p> <p>http://www.isdscotland.org/Products-and-Services/eDRIS/</p> <p>The service includes a research coordinator, help with study design/feasibility, advise on coding, terminology etc.</p>	
Contact details	Linkage	Example Publications	
<p>There is no single contact point for ISD administrative data research enquiries.</p> <p>Researchers should contact eDRIS for help with access and use of health data</p>	<p>Datasets can be linked to each other using Community Health Index (CHI) numbers</p>	<ul style="list-style-type: none"> • Boyle P et al., <i>Cohort profile: The Scottish Longitudinal Study (SLS)</i> Int J Epidemiol 2009 Apr; 38(2):385-92 • Pavis S and Morris AD., <i>Unleashing the power of administrative health data: the Scottish model</i> Public Health Res Pract. 2015 Sep 30;25(4):e2541541 	

10 Screening datasets

10.1 UK Cancer Screening

Name and Description	Type of Data	Coverage	Geography	Duration
NHS Screening data are available for: Breast Cancer Screening Cervical Cancer Screening Bowel Cancer Screening	Screening in primary care	Women who consent to participate living in England, Scotland, Wales and N. Ireland	UK	Varies by dataset
Context of data collected	Volume	Events for data collected	Granularity	Coding
Breast cancer screening programme: women between ages 50 and 70 are invited for regular breast screening (every 3 years) under a national programme Cervical Screening: women between the ages of 25 and 49 are invited for routine screening every 3 years and those aged 50 to 64 are invited every 5 years	Varies by dataset – approx. 70% uptake for breast cancer screening (approx. 2.11 million in 2014-5)	<ul style="list-style-type: none"> Breast: Invitations for breast screening, coverage, uptake of invitations, outcomes of screening, cancers detected Cervical: call and recall, outcomes of screening, referrals to colposcopy clinic 	Per registered patient	ICD-10
Extraction/Data collection	Consent	Access		
Women are invited by the NHS breast screening programme, referred by GP or self-referral	Patient opt in	<p>English data can be downloaded free of charge from www.hscic.gov.uk</p> <ul style="list-style-type: none"> English data from: Breast: http://content.digital.nhs.uk/article/2021/Website-Search?productid=24457&q=cancer+screening&sort=Relevance&size=10&page=1#top Cervical: http://content.digital.nhs.uk/article/2021/Website-Search?productid=23523&q=cancer+screening&sort=Relevance&size=10&page=1#top Bowel cancer screening data for England does not appear to be readily available to researchers Scottish Data from: http://www.isdscotland.org/Health-Topics/Cancer/Breast-Screening/ http://www.isdscotland.org/Health-Topics/Cancer/Cervical-Screening/ http://www.isdscotland.org/Health-Topics/Cancer/Bowel-Screening/ 		

	<ul style="list-style-type: none"> Welsh Data from: http://www.breasttestwales.wales.nhs.uk/reports-1 http://www.cervicalscreeningwales.wales.nhs.uk/statistical-reports N. Irish data from: http://www.cancerscreening.hscni.net
Contact details	Example Publications
See individual website pages	<ul style="list-style-type: none"> Massat N et al., <i>Variation in cervical and breast screening coverage in England: a cross-sectional analysis to characterise districts with atypical behaviour</i> BMJ Open 2015 Jul 24;5(7):e007735

11 Useful websites

Website name	Address	Description
Scottish Primary Care Information Resource (SPIRE)	www.spire.scot.nhs.uk	Due to be launched Spring/Summer 2017 The SPIRE project is a collaboration between the Scottish Government and NHS National Services Scotland (NSS) The aim of SPIRE is to provide a single national system to extract data from General Practice clinical IT systems in Scotland. SPIRE will analyse and report on the data extracted for specific and approved purposes
UK data service	www.ukdataservice.ac.uk	Collection of social, economic and population data resources e.g. health survey for England can be accessed through this website
Health and Social Care Information Centre (HSCIC)	www.hscic.gov.uk	National provider of information, data and IT systems for health and social care Data can be downloaded free of charge e.g. data on breast cancer screening
Information Services Division (ISD) Scotland	www.isdscotland.org	ISD Scotland is part of NHS National Services Scotland. ISD provides health information, health intelligence, statistical services and advise that supports the NHS. Scottish health and administrative data can be accessed through this website
Health in Wales	www.wales.nhs.uk/statisticsanddata	Access Welsh health, social and administrative data through this website
Department of Health, Northern Ireland	https://www.health-ni.gov.uk/topics/doh-statistics-and-research	Access Northern Irish health and social care surveys and statistics