

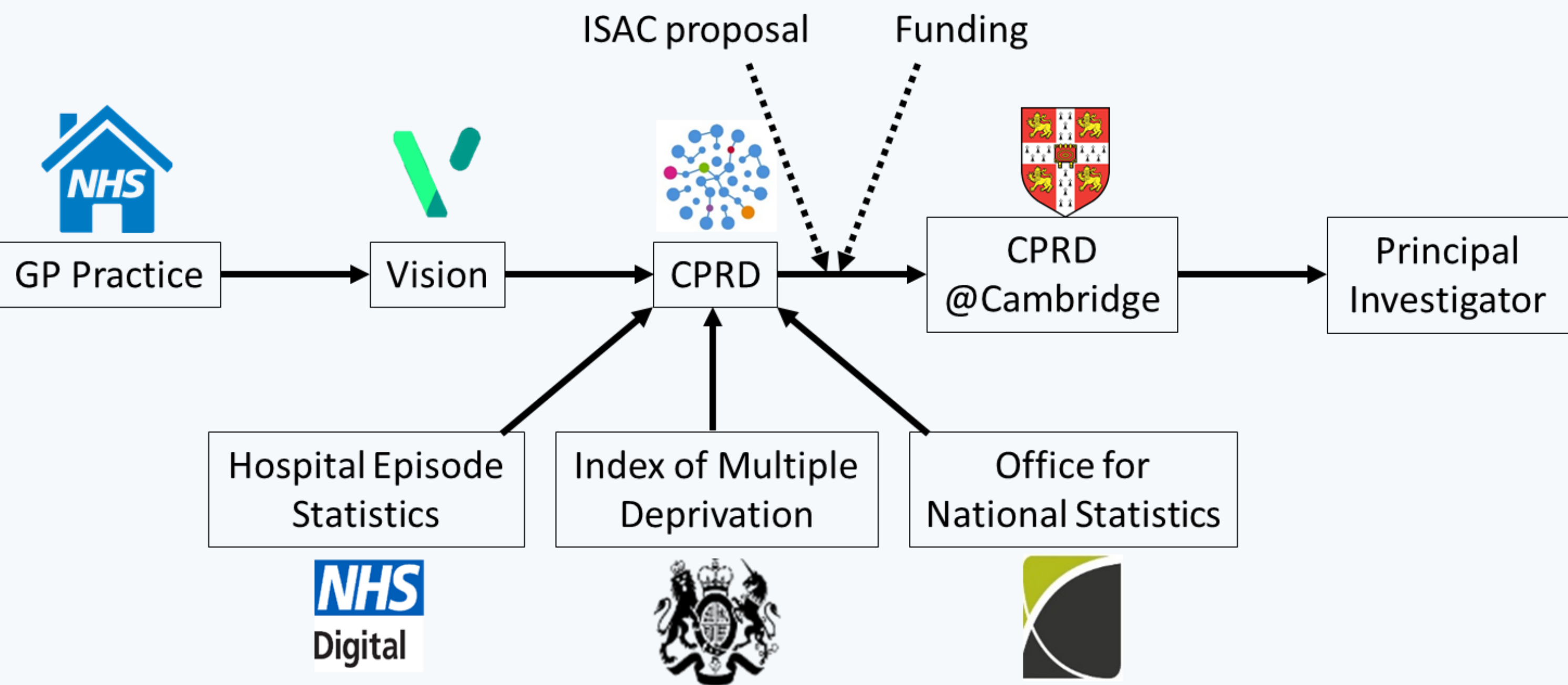
Improving the Identification of Patients in Electronic Health Record Research

Amelia Harshfield* ⁽¹⁾, James Brimicombe ⁽¹⁾, Anna Cassell ⁽¹⁾, Jorge Browne ⁽²⁾, Lucy Gerza⁽¹⁾, Rupert Payne⁽³⁾, Duncan Edwards⁽¹⁾

(1) Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, UK.
(2) Department of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile.
(3) Centre for Academic Primary Care, University of Bristol, Bristol, UK.

Electronic Health Records and the Clinical Practice Research Datalink

Electronic Health Records (EHRs) are routinely collected data, and a digitally stored version of a patient’s paper record. They can contain both coded and free-text data, but epidemiological, primary care, and health services research studies primarily make use of the coded data. It provides a robust means of identifying patients with specific clinical conditions; however, it is usually not collected with research in mind, so caution must be exerted when interpreting the results.



The Clinical Practice Research Datalink (CPRD, www.cprd.com) can be considered one form of an EHR. One of the biggest clinical datasets in the country, it contains approximately 8% of the UK population’s primary care anonymised medical records. A relational database, it includes various information on patients, staff, diagnoses, consultations, prescriptions, referrals, and so on, which are uploaded electronically monthly to CPRD via the computerised Vision practice system. At Cambridge, we can make use of this database by contacting the CPRD@Cambridge team (www.phpc.cam.ac.uk/pcu/cprd_cam) who will be able to provide datasets to researchers who wish to undertake projects using it, on condition that the Independent Scientific Advisory Committee (ISAC) has approved the study proposal and funding is readily available.

Using Code Lists with Electronic Health Records

- ❖ Medical or product codes are building blocks in CPRD for meaningfully defining medical conditions. Medical codes are matched on Read codes, a system set up by James Read in the 1980’s which includes diagnostic and administrative patient data. Product codes are matched on British National Formulary (BNF) codes, which capture prescriptions.
- ❖ Code lists have become a norm when dealing with big routine health data. They are used across national and international electronic health records, of which, examples in the UK include the Clinical Practice Research Datalink (CPRD), The Health Improvement Network (THIN) database, Clinical research using Linked Bespoke studies and Electronic health Records (CALIBER), and the Secure Anonymised Information Linkage (SAIL) databank.
- ❖ The University of Manchester has a repository of code lists found on clinicalcodes.org, that is “an online clinical codes repository to improve validity and reproducibility of medical database research”.
- ❖ Medical and product codes can found in the different tables of CPRD.

Medical Code	Read Code	Read Term	Clinical Events	Referral Events	Test Events
9897	90L..00	Diabetes monitoring admin.	2235926	8957	0
3550	66A..00	Diabetic monitoring	2197902	16054	51
6125	66AS.00	Diabetic annual review	1888891	1830	1
13194	90L4.00	Diabetes monitoring 1st letter	907426	6961	0
711	C10..00	Diabetes mellitus	811572	38887	4
5234	6872	Diabetes mellitus screen	547748	364	16
12506	66AP.00	Diabetes: practice programme	378104	117	1
11471	883L.00	Diabetes medication review	259477	260	0
101177	66At.00	Diabetic dietary review	258050	45	0
1684	66A4.00	Diabetic on oral treatment	229050	445	3
13195	90L5.00	Diabetes monitoring 2nd letter	210980	1329	0
18311	68A7.00	Diabetic retinopathy screening	190645	9062	674410
95994	66Aq.00	Diabetic foot screen	173263	139	0
10977	66Ac.00	Diabetic peripheral neuropathy screening	160193	130	28
13067	66AZ.00	Diabetic monitoring NOS	158284	7493	763
8836	66AR.00	Diabetes management plan given	153213	3945	0
2378	66AJ.00	Diabetic - poor control	134756	8901	26
7563	66A3.00	Diabetic on diet only	106579	568	2
28873	66AI.00	Diabetic 6 month review	102666	195	0
12030	90L6.00	Diabetes monitoring 3rd letter	99686	644	0
			97750	8034	84

- Advantages of code lists**
- ❖ Maintaining consistency of research.
 - ❖ Saves time on data management.
 - ❖ Enables focus on small demographic populations in observational studies and trials.
 - ❖ Usable with minimal changes in other UK EHRs.
- Challenges of using code lists**
- ❖ Data can be ‘messy’ or incomplete, leading to quality issues.
 - ❖ Determining the level of accuracy needed for comprehensible results – trade-off between sensitivity and specificity.
 - ❖ Can find discrepancies between primary and secondary linked data.
 - ❖ Free-text not readily available in CPRD to help interpret the coding used.

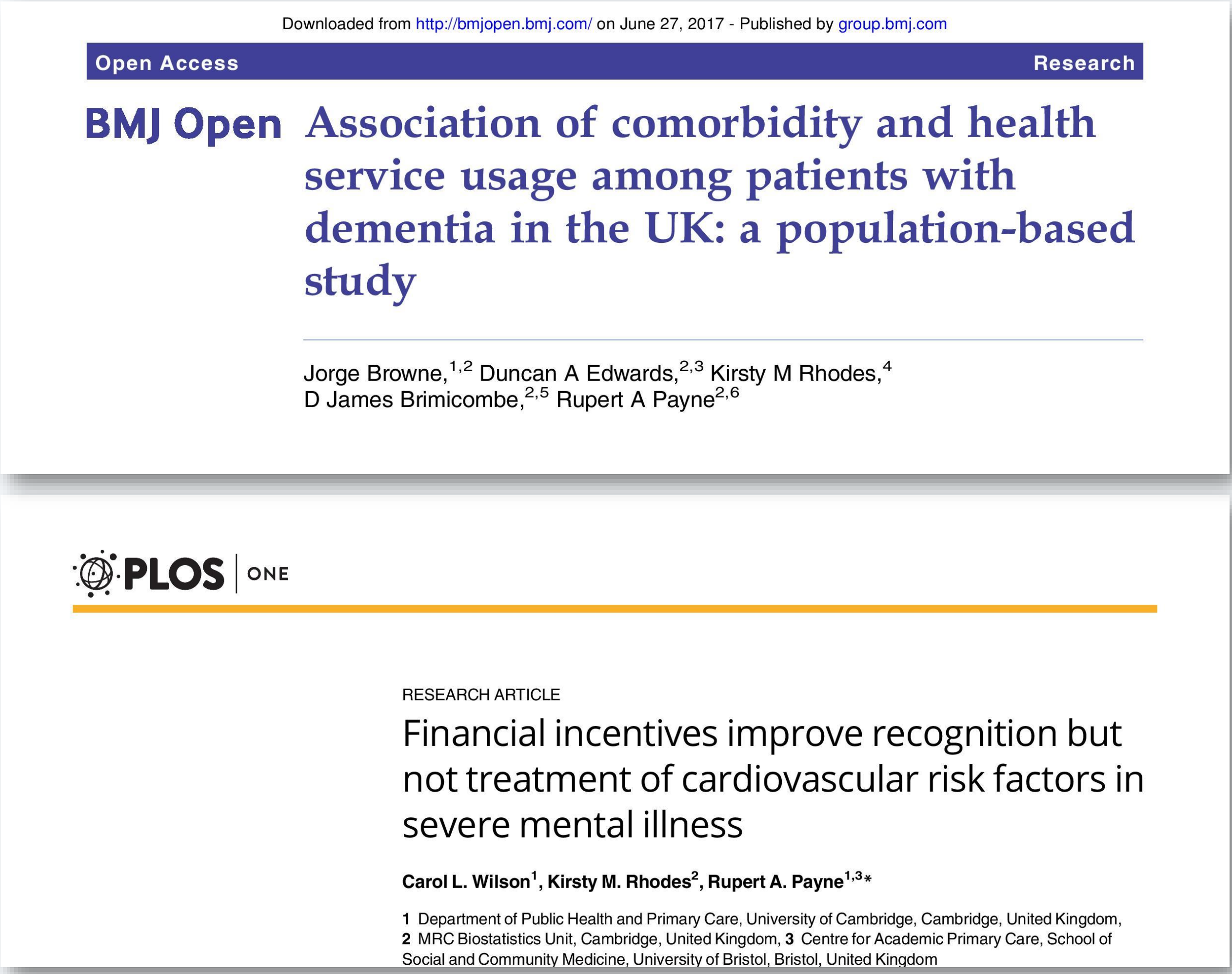
CPRD@Cambridge Code Lists

- ❖ We produced a list of 37 conditions, originally based on work by Barnett (Lancet 2012; 380:37-42).
- ❖ Our diagnostic (medical) code lists were based on existing lists, expert opinion, and on additional terms identified from the relevant Read code hierarchy. In some cases, administrative codes were included.
- ❖ Our prescription (product) code lists were also based on existing lists and expert opinion, but included text word searches for relevant branded and generic products within the corresponding sections of the BNF.
- ❖ Each condition is made up of at least one of the two types of code list; however, a subset of conditions use two lists in conjunction.
- ❖ Cross-sectional prevalence checks were run based on the April 2016 CPRD dataset population.

Condition	Prevalence	Usage
Alcohol problems	1.80%	Read code ever recorded
Anorexia or bulimia	0.50%	Read code ever recorded
Anxiety & other neurotic, stress related & somatoform disorders	17.00%	Read code in last 12 months OR = 4 anxiolytic/hypnotic prescriptions in last 12 months
Asthma (currently treated)	3.70%	Read code ever recorded AND any prescription in the last 12 months
Atrial fibrillation	2.20%	Read code ever recorded
Blindness and low vision	1.00%	Read code ever recorded
Bronchiectasis	0.40%	Read code ever recorded
Cancer - diagnosis in last five years	1.80%	Read code first recorded in last 5 years
Chronic liver disease and viral hepatitis	0.40%	Read code ever recorded
Chronic kidney disease	3.00%	Best of last 2 GFRs less than 60 mL/min OR Read code ever recorded
Chronic Sinusitis	2.50%	Read code ever recorded
COPD	2.20%	Read code ever recorded
Constipation (treated)	2.20%	4 or more laxative prescriptions in last year
Coronary heart disease	4.30%	Read code ever recorded
Dementia	0.70%	Read code ever recorded
Depression	10.30%	Read code recorded in last 12 months OR >=4 anti-depressant prescriptions (excluding low dose tricyclics) in last 12 months
Diabetes	5.90%	Read code ever recorded
Diverticular disease of intestine	2.80%	Read code ever recorded
Epilepsy (currently treated)	0.80%	Read coded ever recorded AND antiepileptic prescription in last 12 months
Hearing loss	9.50%	Read code ever recorded
Heart failure	1.00%	Read code ever recorded
Hypertension	18.80%	Read code ever recorded
Inflammatory bowel disease	0.50%	Read code ever recorded
Irritable bowel syndrome	7.90%	Read code ever recorded OR =4 prescription only medicine antispasmodic prescription
Learning disability	0.40%	Read code ever recorded
Migraine	0.40%	4 or more prescription only medicine anti-migraine prescriptions in last year
Multiple sclerosis	0.30%	Read code ever recorded
Painful condition	10.10%	>=4 POM analgesics in last 12 months OR >=4 specified anti-epileptics in the absence of an epilepsy Read code in last 12 months
Parkinson's disease	0.30%	Read code ever recorded
Prostate disorders	5.70%	Read code ever recorded
Peripheral vascular disease	0.9%	Read code ever recorded
Psoriasis or eczema	1.00%	Read code ever recorded AND =4 related prescriptions in last 12 month (excluding simple emollients)
Psychoactive substance misuse	1.20%	Read code ever recorded
Rheumatoid arthritis, other inflammatory, polyarthropathies & systemic connective tissue disorders	2.20%	Read code ever recorded
Schizophrenia (and related non-organic psychosis) or bipolar disorder	1.00%	Read code ever recorded/recorded in last 12 months (code dependent) OR Lithium
Stroke & transient ischaemic attack	2.20%	Read code ever recorded
Thyroid disorders	4.70%	Read code ever recorded

Example Projects

- ❖ A number of projects have made use of the code lists, for example:
 - Investigating the association of morbidity and health service usage among patients with dementia in the UK.
 - Evaluating the effect of financial incentives in improving the recognition and treatment of cardiovascular risk factors in those with severe mental illness.



The Value of Our Code Lists

- ❖ This is the first time a set of code lists have been systematically and methodically been produced and published, lending itself to:
 - ✓ **Transparency**: our code lists, along with their methodological development, have been published on an online repository which anybody can access. They can be easily downloaded and used.
 - ✓ **Reproducibility**: they can be used as is or adapted where appropriate. Additionally, future research can be verified based on established code lists.
 - ✓ **Efficiency**: they minimise the amount of work needed to create a future code list from scratch and can be built upon.

