13

Electronic Health Record Databases

Daniel B. Horton[1], Harshvinder Bhullar[2], Lucy Carty[3], Francesca Cunningham[4], Alexis Ogdie[5], Janet Sultana[6], and Gianluca Trifirò[6]

[1] *Department of Pediatrics, Rutgers Robert Wood Johnson Medical School, Rutgers Center for Pharmacoepidemiology and Treatment Science, New Brunswick, NJ, USA*
[2] *Real-World Insights, IQVIA, London, UK*
[3] *Clinical Practice Research Datalink, Medicines and Healthcare products Regulatory Agency, London, UK*
[4] *Pharmacy Benefits Management, and Center for Medication Safety, US Department of Veterans Affairs, Hines, IL, USA*
[5] *University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA*
[6] *Department of Biomedical and Dental Sciences and Morphofunctional Imaging, University of Messina, Messina, Italy*

Databases that contain health information can be divided into two broad categories: those that collect information for administrative purposes, such as filing claims for payment (administrative claims databases), and those that serve as the patient's medical record (electronic health record [EHR] databases), which physicians use to track health information about their patients. Administrative claims databases are maintained for billing or administrative purposes rather than for the actual provision of patient care. In contrast, data from EHR databases are thought more likely to be clinically accurate because they are collected for patient care, recorded by clinicians (versus coders), and reflect information that may not relate to billing. Unlike administrative databases, EHR databases are more likely to capture important health information about patients, such as symptoms of illness, historical data, family history, smoking and alcohol use, vital signs (e.g., body mass index [BMI]), and laboratory data [1]. Of note, we use the term *electronic health record* databases to encompass their use in the provision of clinical care (that is, as medical records) as well as the interoperability of these electronic systems across broad healthcare networks, serving various nonclinical functions, such as administration, billing, and research [2].

Despite their many advantages, EHR databases have certain limitations. Some EHR databases, like that of the Veterans Affairs (VA) and other US EHRs, may not capture diagnoses and treatments from out-of-system care. Other EHR databases, particularly European primary care databases, lack information from secondary care settings (e.g., hospitals and specialists), and linkage to secondary care datasets is not available for some databases. While EHR databases are thought to have greater clinical accuracy in recorded diagnoses, one cannot presume the validity of diagnostic codes without formal validation. EHR databases usually contain data on prescribed outpatient drugs, but many databases lack information on drug dispensing or inpatient medications. In addition, there may be a high proportion of

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Table 13.1 Overview of EHR databases.

	Arianna	BIFAP	CPRD	LPD Italy	IPCI
Country or countries (and region)	Italy (Caserta)	Spain	United Kingdom	Italy	Netherlands
Year data collection initiated	2000	2001	1987	1998	1989
Number of patients and amount of follow-up time*	0.6M patients, 3.4M person-years	7.9M patients; 49.7M person-years	30M patients (>15M research standard patients; 10M actively followed patients); >100M person-years	1.6M patients; 19.2M person-years	2.4M patients; >12.1M person-years [166]
Sex distribution*	Female: 0.31M (51.0%) Male: 0.28M (49.0%)	Female: 4.1M (52.3%) Male: 3.8M (47.7%)	Female: 49.3% Male: 50.7% [18]	Female: 0.83M (51.8%) Male: 0.77M (48.2%)	Female: 1.2M (51.2%) Male: 1.2M (48.8%)
Age distribution*	0–10: 0.01M (1.3%) 11–20: 0.05M (8.6%) 21–44: 0.22M (36.2%) 45–64: 0.18M (30.4%) 65–84: 0.11M (18.9%) ≥65: 0.03M (4.5%) Unknown: 0.03M (4.3%)	0–10: 1.2M (15.1%) 11–20: 0.72M (9.1%) 21–44: 2.9M (36.6%) 45–64: 1.7M (21.6%) 65–84: 1.2M (15.0%) ≥85: 0.20M (2.6%)	<18: 20.2% 18–64: 61.8% ≥65: 18.1% [18]	0–14: 0 (0%) 15–24: 0.08M (9.9%) 25–44: 0.48M (30.3%) 45–64: 0.54M (33.9%) 65–84: 0.35M (21.9%) ≥85: 0.06M (4.0%)	0–9: 0.26M (10.8%) 10–19: 0.27M (11.3%) 20–39: 0.64M (26.5%) 40–59: 0.65M (23.9%) 60–79: 0.46M (19.1%) ≥80: 0.13M (5.5%)
Race and ethnicity distribution	Not available	Mostly not available; some PCPs may have recorded race and ethnicity in free text	Available for 27% of patients in CPRD and 79% of inpatients in HES [167]. Known race/ethnicity distribution: White: 87% South Asian: 6% Black: 4% Mixed: 1% Other: 2%	Not available	Not available
Number of physicians or practices included*	300 GPs [34]	4910 GPs, 842 pediatricians	851 practices (actively contributing)	800 GPs [168]	Approximately 600 GPs from 200 practices [169]
Diagnostic coding system	ICD-9	ICD-9, ICPC	Read, ICD-10 (HES)	ICD-9	ICPC
Drug coding system	ATC, NDC	ATC	Gemscript	ATC	ATC

*Numbers updated through 2017.

IQVIA DA	Pedianet	SIDIAP	THIN	VA
Germany, France, United Kingdom	Italy	Spain (Catalonia)	United Kingdom	United States
1992	2000	2006	2002	1997
Germany: 34M patients (including 17.2M German specialty patients); 54.5M person-years France: 10.5M patients; 6.0M person-years [22] UK: 4.2M patients; 17.9M person-years	0.4M pediatric patients (0.2M actively followed); 1.8M person-years	5.6M patients; 5.7M person-years	18M patients (3.1M actively followed patients); >90M person-years	14.5M patients; 168M person-years
Germany Female: 19M (41%) Male: 15M (32%) Unknown: 12.5M (27%) France Female: 5.5M (52%) Male: 5M (47%) Unknown: 0.06M (1%)	Active patients: Females: 0.10M (48.2%) Males: 0.11M (51.8%)	Female: 2.9M (50.7%) Male: 2.8M (49.2%)	Female: 9.4M (52.2%) Male: 8.6M (47.8%)	Female: 1.9M (13.1%) Male: 12.5M (86.4%) Unknown: 0.07M (0.5%)
0–9: 1.5M (4.4%) 10–19: 2.6M (7.7%) 20–39: 7.9M (23.2%) 40–59: 9.5M (28.0%) 60–79: 8.2M (24.1%) ≥80: 4.3M (12.6%)	Active patients: 0–4: 0.04M (19.6%) 5–9: 0.06M (28.6) 10–16: 0.11M (51.8%)	0–10: 0.63M (11.2%) 11–20: 0.55M (9.7%) 21–44: 1.9M (33.3%) 45–64: 1.5M (27.1%) 65–84: 0.88M (15.6%) ≥85: 0.17M (3.1%)	Active patients: 0–10: 0.39M (12%) 11–20: 0.35M (11%) 21–44: 1M (32%) 45–64: 0.85M (27%) 65–84: 0.50M (16%) ≥85: 0.08M (2%)	<21: 0.24M (1.7%) 21–44: 2.4M (16.6%) 45–64: 3.5M (24.0%) 65–84: 5.3M (36.8%) ≥85: 2.4M (16.4%) Unknown: 0.65M (4.5%)
Not available	Not available	Not available	Not available	White 9.0M (62.4%) Black 1.9M (12.9%) Asian 0.24M (1.7%) Hispanic 0.48M (3.4%) Other 0.11M (0.8%) Unknown 2.8M (18.8%)
Germany: 2357 general practices, 2010 specialty practices France: 2091 practices UK: 218 practices [22]	300 family pediatricians	3414 GPs, 853 primary care pediatricians [170]	Over 700 practices	Healthcare professionals: 127 211 Physicians: 23 973
ICD-10, Read (UK)	ICD-9, ICD-10	ICD-10	Read, ICD-10 (HES)	ICD-9, ICD-10, CPT
ATC	ATC, NDC, Italian MINSAN codes	ATC, NDC	Gemscript	VA Drug Classification System [171], NDC

(*Continued*)

Table 13.1 (Continued)

	Arianna	BIFAP	CPRD	LPD Italy	IPCI
Software used	SANI.A.R.P.	Various, mainly OMI-AP [37]	Vision and EMIS	Millewin	Various
Quality checks, standards, and feedback to clinicians	Arianna conducts periodic data checks. GPs whose recorded data do not meet preestablished standards are excluded from research [34]	BIFAP performs extensive quality and validity checks of raw data from individual autonomous regions, including physician- and patient-level data. Patient data that are inconsistent or fail quality checks are excluded. BIFAP provides feedback to participating GPs and pediatricians by comparing their patients' registrations, disease characteristics, and prescribing indicators with those of other clinicians within BIFAP	CPRD performs permanent, ongoing quality checks of data from all practices. Patients with noncontiguous follow-up or poor data recording are excluded. Remaining patients are flagged as acceptable for use in research. See text for description of Up-to-Standard dates	IQVIA performs periodic quality checks based on coding accuracy, concordance between GP-specific and national prevalence of selected diseases, and mortality rates [168]. Data from GPs that do not meet set standards are excluded [25]	IPCI evaluates each GP practice for data quality based on several indicators. Data from practices below a preestablished quality threshold are excluded from research. GPs are not permitted to keep paper-based records to improve data quality [172]
Data access and approval	Arianna is available through collaboration with either the Local Health Unit of Caserta or academic institutions with data access, such as the University of Campania or the University of Messina. Researchers must first notify their local ethics committee before using the data. Full ethics committee evaluation is not needed (anonymized data, no direct patient interaction)	BIFAP is available to affiliated and other noncommercial researchers. Investigators must receive approval from the BIFAP scientific committee	CPRD (www.cprd.com) licenses online access to the database. Researchers can download CPRD data using a secure file transfer protocol. CPRD receives annual regulatory ethics approval to supply anonymized linked data for public health research. All research requests to access data held by CPRD are reviewed by the Independent Scientific Advisory Committee	Within Italy, LPD Italy (www.healthsearch.it/?lang=en) is available through collaboration with the Italian College of General Practitioners or with IQVIA (www.iqvia.com)	Access to IPCI (www.ipci.nl/Framework/Framework.php) is provided through collaboration with the Erasmus Medical Centre. Protocols for studies using IPCI data must be approved by the IPCI ethics committee [174]

IQVIA DA	Pedianet	SIDIAP	THIN	VA
Various	Junior Bit	e-CAPTM	Vision	Various
IQVIA checks all data for quality standards and plausibility. IQVIA gives all physicians monthly feedback reports showing their prescription patterns and those of colleagues within the IQVIA panel and within their specialty group. Data from DA Germany are also checked annually by the German Medical Association	Data quality in Pedianet is evaluated for every study conducted, either by a central database unit or by researchers. Quality checks include validation of ICD-9-based diagnoses in the clinical chart and free text	SIDIAP performs systematic quality checks to harmonize data and identify duplicate patients, logical errors, and implausible values and inconsistent units within laboratory data. SIDIAP initially used the Registry Quality Standard score to assess data quality [173], but this has been discontinued	IQVIA performs ongoing consistency and integrity checks on all THIN data. See text for description of Acceptable Mortality Reporting	All VA data are updated regularly and checked for quality. Drug data are updated daily with quality checks
IQVIA (www.iqvia.com) administers DA France or Germany with various options: researchers can buy the software and data with monthly updates, preprocessed datasets, or data analyzed directly by IQVIA. Approval for IQVIA DA France and Germany requires only local IRB approval	Pedianet data (pedianet.it/en/about) can be obtained by collaboration with Pedianet-affiliated epidemiologists	Researchers must pay a fee and sign an agreement to obtain data from SIDIAP. Commercial organizations may not use SIDIAP data directly but can contract the core SIDIAP research team to conduct studies with input from the scientific and ethical committee	IQVIA (www.iqvia.com) makes THIN data available in a few forms: a sublicense for the whole research-formatted dataset, a data subset, or a preprocessed dataset with some data manipulation. THIN studies require approval by the Scientific Review Committee of independent researchers. If additional information will be collected, ethics approval is required from the NHS Multi-Centre Research Ethics Committee	Access to VA data is limited to researchers employed by the VA or with VA appointments, and their collaborators. Approval by the local or central VA IRB is required

ATC, Anatomical Therapeutic Chemical; CPT, Current Procedural Terminology; GPs, general practitioners; HES, Hospital Episode Statistics; ICD, International Classification of Diseases; ICPC, Classification of Primary Care; M, million; NDC, National Drug Code; NHS, National Health Service; PCP, primary care professional.

Table 13.2 Selected variables in EHR databases available for epidemiologic research.

	Arianna	BIFAP	CPRD	LPD Italy	IPCI
Healthcare professional demographics	Arianna provides physicians' age, sex, years since graduation	BIFAP provides geographic area of GP but not GPs' demographic data. One cannot determine whether nurse or doctor entered data	CPRD reports if nurse or doctor entered data	One can identify GPs' geographic region but not demographics	GPs' demographics are not available
Types of physicians	GPs	Mainly GPs but also other members of the primary care team, such as pediatricians and nurses	GPs	GPs	GPs
Practice and patient demographics	Practice: Location Patient: DOB, sex, healthcare exemption (based on salary and disability)	Practice: Number of patients registered with GP; number of persons registered in practice available upon request Patient: DOB, sex	Practice: Region, practice size, practice-level SES (Index of Multiple Deprivation and Townsend scores, ~60%), date of last registration, Up-to-Standard date (see text) Patient: YOB for adults, month and YOB for children; sex, ethnicity (~25% recorded; also available via census data), census-based socioeconomic class; patient status (active, died, transferred out)	Practice: Location Patient: DOB, sex healthcare exemption (based on salary and disability)	Practice: Number of employees may be available for some practices Patient: DOB, sex
Vital signs and social history	Height, weight, BMI, smoking, alcohol use available for 25% persons aged ≥65 (2013–present); BP available for some patients (2016–present)	Weight, BMI, BP, smoking, and alcohol consumption	Height, weight, BP, and BMI recorded but may be biased towards patients with a more relevant need for these measurements; smoking (83–93%) [175,176], obesity (61–79%) [175–177], alcohol (~80%) [175,178]	BMI [179], BP [172], smoking, alcohol intake [180]	BP, weight, BMI, and smoking [181] available but recorded only when GPs consider them relevant

IQVIA DA	Pedianet	SIDIAP	THIN	VA
IQVIA DA provides physicians' age, sex, and years in practice	Pedianet provides pediatricians' age, sex, and city of clinic. More detailed information (e.g., years since graduation) is available by request	SIDIAP provides age, sex, type of primary care professional, and performance indicators (quality of care, quality of prescriptions, and quality of diagnosis)	THIN reports if nurse or doctor entered data	One can identify the doctor, nurse, or pharmacist who entered prescription data
Mainly GPs; IQVIA DA Germany and France also include specialists (e.g., cardiologists, dermatologists)	Family pediatricians	Health professionals working in primary care: GPs, pediatricians, dentists, nurses, midwives	GPs	Mainly PCPs but also physician specialists (e.g., cardiologists) and other clinicians (e.g., nurse practitioners, clinical pharmacy specialists)
Practice: Region, community size, patients per practice, number of doctors, number of employees, type (e.g., GP vs specialty) Patient: Age, sex, health insurance status (e.g., private, statutory), medical insurance company, region, town size (>100 000 vs <100 000)	Practice: Region, patients per practice Patient: YOB, age, sex, region of residence, nationality, information about parents (e.g., nationality, habits, blood group, mother's educational level, socioeconomic level)	Practice: Location, urban/rural, number of patients, deprivation index (MEDEA) Patient: DOB, sex, country of origin	Practice Region, number of patients, computerization date, Vision date, Acceptable Mortality Reporting (see text) Patient: YOB for adults, month and YOB for children; patient-level, location-based socioeconomic status (Townsend deprivation scores, 95% recording), region, ethnicity, (20% recording), patient status (active, died, transferred out)	Practice: Region, facility, type of facility (medical center clinics vs community-based outpatient clinics), facility's level of complexity Patient: DOB, sex, race, ethnicity, zip code
BMI (~40% [182]); smoking and alcohol recording unknown	Gestational age, birth weight, birth height, neonatal jaundice; growth measurements (e.g., height, weight); parental smoking	BP, BMI, smoking, alcohol intake, Framingham score. Pediatric screening data (height, weight, head circumference, pubertal development)	Height, weight, BP, and BMI recorded but may be biased towards patients with a more relevant need for these measurements; smoking (86–94%) [86,183–185], obesity (73–83%) [185], alcohol intake (75–85%) [185]	BP, HR, height, weight, SES, education, marital status, smoking history (>90%)

(*Continued*)

Table 13.2 (Continued)

	Arianna	BIFAP	CPRD	LPD Italy	IPCI
Referrals, procedures, results of investigations	Laboratory test results for ~25%; linkage to hospital discharge data, referral data, and orders for diagnostic tests	PCPs' referrals to specialists and hospitals; results from referrals may be recorded in coded fields or as free text; self-referrals (less common) not available	Detailed information on referrals, procedures, and laboratory tests available for approximately 75% of all patients through linkage to HES	Referral data and orders for diagnostic tests available for all patients	Often not available; letters from hospitals to GPs with test results may be available for some practices; test results may be manually recorded in free-text notes
Type of drug data	Drugs prescribed in community setting; drug dispensing by linkage to claims	Drugs prescribed and dispensed in community setting; vaccine data available	Drugs prescribed in primary care; some OTC drug data available (see text); vaccine data available	Drugs prescribed in community setting; vaccine data available	Drugs prescribed in community setting; vaccine data available
Available drug information	Drug ATC code, NDC (with brand, formulation, units), indication for use	Drug name, active substance, number of prescribed packages, duration, prescribed daily dose, strength, indication for use	Drug name, route, strength, frequency, duration; immunizations including batch; cost of therapy upon request	Drug name, route, dose, frequency, duration, cost of therapy	Drug name, quantity, strength, dose [187]
Health care utilization	GP visits; hospital discharge letters, referrals to specialists, admission to ED available by linking with claims data	GP visits; referrals by GP to secondary care and ED; hospital admissions available if patients referred to GPs after discharge	GP visits, hospitalizations, and consultant visits; links to HES provide detailed ward-level resource utilization (England only)	GP visits, hospital discharge letters, referrals to specialists	GP visits; other data generally not available unless hospital discharge letters sent to GP
Identification of pregnancy and families	ICD-9 codes for pregnancy or birth by linkage to claims data; cannot identify families	ICD-9/ICPC codes for pregnancy; cannot identify families	Pregnancy and pregnancy outcomes, family/ household identification number; mother–baby link via family/ household number and algorithm	Not available	Some birth-related data available through hospital discharge letters; cannot identify families

IQVIA DA	Pedianet	SIDIAP	THIN	VA
HbA1c, blood glucose, cholesterol, LDL, HDL available; other test results variably available but can be requested from paper files	Apgar scores, laboratory and imaging tests ordered and reasons for request; test results not always available	Laboratory tests (date, results), diagnostic and imaging referrals; spirometry; referrals for therapeutic procedures; referrals to secondary and tertiary care [186] (date, reason of referral [ICD-10], specialty referred)	Electronic referrals available; older referrals may be in paper files; most outpatient laboratory results available	Provider referrals for specialists available; all laboratory results available but must be standardized
Drugs prescribed	Drugs prescribed and dispensed in community setting, including drugs not covered by healthcare system; inpatient drug data available if reported to pediatrician; noncompulsory vaccine data available, remaining vaccine data identified via linked claims	Drugs prescribed and dispensed in community setting for drugs covered by the national healthcare system; vaccine data available	Drugs prescribed in primary care; vaccine data available	Drugs prescribed and dispensed in outpatient and inpatient settings; vaccine data available
Drug name, route, dose, frequency, duration, cost of therapy	Drug name, ATC code, indication for use, Italian MINSAN code, NDC (with brand, formulation, units), number of prescribed packages, dose (not available for 30%)	ATC code, NDC, indication for use, profession of prescriber; prescribing data only: start and end date, drug units per day; dispensing data only: units per package, number of packages per month, month of drug dispensation	Drug name, route, strength, frequency, duration; immunizations including batch; linkage available to cost of therapy	Drug name, route, strength, dose, frequency, quantity, duration; cost of therapy
GP visits, hospitalizations, sick leave	Pediatrician visits, ED or hospital admission if referred by pediatrician	PCP visits, referrals to secondary and tertiary care, sick leave (date, length, ICD-10), hospital discharge	GP visits, hospitalizations entered by GP, sick leave (if issued by GP); links to HES provides detailed ward-level resource utilization (England only)	Outpatient visits, ED visits, hospitalization (including medical surgical, and intensive care units), community living center (VA nursing home)
Pregnancy variable, gynecologist records; family data incomplete	May identify siblings	Pregnancy and pregnancy outcomes; mother–baby link available	Pregnancy and pregnancy outcomes; mother–baby link via family/household number and algorithm	ICD-9/ICD-10 codes for pregnancy

(*Continued*)

Table 13.2 (Continued)

	Arianna	BIFAP	CPRD	LPD Italy	IPCI
Identification of death and cause of death	Date of death by linkage to claim data	Date of death; cause of death not available consistently	Date and cause of death available via CPRD data and linkage to Office for National Statistics	Date of death	Date of death; cause of death available via free text
Additional data, such as consult records, free text, or paper files	No free text or letters available	Anonymized free-text notes by GPs are available	Hospital discharge summaries, consultant letters; no free text available	No free text or letters available	Free text available on request
Questionnaires and investigator-initiated outcome validation	Not possible to administer questionnaires	Questionnaires can be given to GPs	Questionnaires can be given to GPs and patients; response rates from three recent studies were ~90% [188] (and CPRD internal data)	Not possible to administer questionnaires	Possible to administer questionnaires but response rates usually low
Settings and types of missing data	Inpatient data (except via discharge forms with main diagnoses), laboratory results for 75%, OTC drugs, vaccines	Inpatient data, OTC drugs (few OTC drugs in Spanish national healthcare system)	Prescriptions in secondary care, OTC drugs (exceptions in text), drug dispensing, adherence	Inpatient data, OTC drugs, drug dispensing, pediatric clinical and prescribing data (any setting)	Inpatient and specialist data, OTC drugs, drug dispensing; linkage available to Dutch PHARMO database with dispensing data

IQVIA DA	Pedianet	SIDIAP	THIN	VA
Date and cause of death seldom recorded	Date and cause of death	Date of death	Death date, sometimes cause of death; death certificates may be accessed for a fee if ethics approval is obtained	Date of death
No free text available	Free text available on request	Hospital discharge for 30% of the SIDIAP patients; other data available by request	Hospital discharge summaries, consultant letters; no free text available	Additional data including consult records and free text available by chart review
Questionnaires available upon request	Patients and families can be contacted for structured or unstructured interviews by phone calls from participating pediatricians	Questionnaires can be given to sample of GPs	Questionnaires can be given to GPs and patients; response rates to paper questionnaires ~90% [51] (and THIN internal data)	Charts may be reviewed for validation
Secondary care records, vaccine data, linkage between patients seen in both primary care and specialty clinics	Inpatient data not available for 60%; OTC drug data (unless reported to pediatrician); adult health data	Inpatient data not available except admission/discharge data for hospitals of the Catalan Health Institute; OTC drugs, indication for drug use, drugs not covered by national health system	Prescriptions in secondary care, OTC drugs (exceptions in text), drug dispensing, adherence	Encounter and drug data from healthcare facilities outside VHA, including for patients taken to nearby hospitals for acute events (e.g., stroke); some inpatient medications housed in floor stock for acute care

BP, blood pressure; DOB, date of birth; ED, emergency department; GPs, general practitioners; HES, Hospital Episode Statistics; ICD, International Classification of Diseases; ICPC, International Classification of Primary Care; NDC, National Drug Code; OTC, over the counter; PCP, primary care professional; SES, socioeconomic status; VHA, Veterans Health Administration; YOB, year of birth.

missing data for some variables of interest, such as disease severity, smoking history, body mass index, and patients' race or occupation.

In this chapter, we focus on selected primary care EHR databases from Europe and a national EHR database for veterans from the United States. Several European databases are available for licensing by investigators in government, academia, and industry, while access to certain European databases and to VA data requires collaboration with affiliated researchers. These databases have been used by epidemiologists, and in particular pharmacoepidemiologists, resulting in thousands of published studies. While there are many similarities among the databases, there are also important differences, which we describe in more detail below (see also Tables 13.1 and 13.2). Of note, we do not discuss other outpatient EHR databases that may have been used in pharmacoepidemiologic research but are less widely utilized or less representative of broader source populations. Claims databases from various countries are covered in Chapter 12, and inpatient EHR databases will be discussed at length in Chapter 14.

Description

Europe and the United Kingdom

Overview of Healthcare Systems and Populations

France, Italy [3], Spain [4], and the UK have universal, government-funded healthcare systems. Germany and the Netherlands require all persons to have medical insurance to cover healthcare [5]. In several of these and other European countries, general practitioners (GPs) act as gatekeepers for medical care. In Italy and Spain, family pediatricians function similarly as gatekeepers of most children's healthcare. Except in certain European countries (e.g., France, Germany), practically the entire population have primary care professionals (GPs or family pediatricians), and the vast majority of these clinicians have EHRs. Where GPs and pediatricians act as gatekeepers of the health system, they not only provide general (primary) medical care but are also involved in or informed of nearly all medical events involving their patients, including referrals to specialists, admission to emergency departments or hospitals, and prescribing of medicines recommended by consulting specialists. Thus, European primary care-based EHR databases capture most of their patients' health information.

Of note, the Italian and Spanish healthcare systems are strongly decentralized [6]. Healthcare services in these countries are managed and provided at the regional level, and their respective EHR databases reflect this regionalization. Notably, while both France and Germany have universal healthcare systems, patients often have additional private insurance, and GPs have less of a gatekeeper role than in other countries. These more open healthcare systems, therefore, make the French and German EHR databases less complete records of patients' health information [7–10].

Overview of Databases

The UK was the setting of the first European EHR database, Clinical Practice Research Datalink® (CPRD®), (previously known as Value Added Medical Products [VAMP] database and then the General Practice Research Database® [GPRD®]). CPRD was established in 1987 as a tool for conducting public health research. The Dutch IPCI database followed shortly thereafter in 1989. Since then, multiple other European health record databases were developed and used for research purposes (see below and Table 13.1).

Clinical Practice Research Datalink is a research service of the UK government, supported by the Medicines and Healthcare products Regulatory Agency (MHRA) and National Institute of Health Research. The Health Improvement Network® (THIN®) was set up in 2002 as a collaboration between software and database companies (respectively, Cegedim and

Epic Database Research Company Ltd, now part of IQVIA; THIN is a Cegedim database). CPRD and THIN collect similar health information from approximately 3–5 million individuals per year seen by GPs in the UK, representing 5–8% of the population [11–19]. The same practices may contribute to both CPRD and THIN, but the proportion of overlap changes over time as new practices join or leave each database.

In a study validating well-established drug–outcome associations from the literature, findings were similar in CPRD practices and non-CPRD practices within THIN [16]. As an example of the overlap of these two databases, one study identified over 60% of individuals initiating a particular drug in both THIN and CPRD [20]. Increasingly, pharmacoepidemiologists are combining information from both databases to increase sample size and improve statistical power and generalizability. Because CPRD and THIN are not mutually exclusive, merging data requires identification and singular inclusion of practices contributing to both databases in a given year. Investigators have developed an algorithm for identifying overlapping practices while maintaining anonymity based on total numbers of patients per practice stratified by gender and birth year [21].

The IQVIA Disease Analyzer databases (DA, previously known as Mediplus®) were set up in France, Germany, and the UK. Because the DA UK database is no longer available, it will not be discussed in detail, but additional information may be found in Tables 13.1 and 13.2. The IQVIA DA databases include anonymized patient records from primary care practices as well as some office-based specialists, including cardiologists, dermatologists, diabetes specialists, gynecologists, neurologists, orthopedists, otolaryngologists, pediatricians, psychiatrists, and urologists [22,23]. Patients who see both general practitioners and specialists have different identity codes in the databases to preserve patient confidentiality, making it challenging to track patients across different settings of care. With nearly 30 million patients (5–7% of the total population), DA Germany is larger than DA France, which has over 10 million patients (16% of the total population).

In Italy, the Health Search Longitudinal Patient Database (LPD Italy from IQVIA) contains data on 1.6 million people from GPs across the nation (2.6% of the population), making it the country's largest EHR database [24–29]. Founded in 1998 by the Italian College of General Practitioners, LPD Italy is now owned by IQVIA. The Arianna database contains EHR data on approximately 600 000 people (60% of inhabitants) in a region of southern Italy [30–34]. The Arianna database is the only Italian EHR database that systematically links to several administrative claims databases and includes drug dispensing and hospital discharge data (see Chapter 12 for other data sources combining claims and EHR data). Of note, Arianna data may be linked to comprehensive geriatric assessments (systematic, multidimensional evaluations of health status covering cognitive and physical function, mobility, disability, social support, etc.) for almost three-quarters of the local elderly population (90 000 since 2014), making it a valuable resource for geriatric research [35]. On the other side of the age spectrum, Pedianet contains data on over 400 000 children throughout Italy since 2000, over half of whom are being actively followed (see Chapter 22 for other data resources for pediatric pharmacoepidemiologic research).

In Spain, Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP) and Sistema de Información para el Desarrollo de la Investigación en Atención Primaria (SIDIAP) are primary care EHR databases that differ in catchment area. Founded in 2000 by the Spanish Agency on Medicines and Medical Devices, BIFAP contains data on 7.9 million individuals from nine of 17 autonomous communities (17% of the Spanish population) [36,37]. SIDIAP was founded in 2006 by the Catalan Institute of Health and the Primary Care Research Institute Jordi Gol. SIDIAP contains data on almost 5.6 million persons from a single

autonomous community, Catalonia (85% of the Catalonian population).

General practitioners across the Netherlands contribute data to Integrated Primary Care Information (IPCI, previously known as Interdisciplinary Processing of Clinical Information). As of 2018, 2.4 million individuals (14% of the Dutch population) are registered in IPCI.

Data Collection and Structure

Practitioners use electronic health records to document a wide range of clinical information about their patients. These data are then electronically extracted for research purposes by specialized software (see Table 13.1), examined for completeness and accuracy by database administrators, and uploaded in anonymized form into the database. The frequency of data extraction and transmission varies among databases. Some EHR databases receive frequent data updates (e.g., real-time contributions in Pedianet, daily in CPRD, three times per week in THIN, monthly in Arianna). Other databases receive new clinical data just 1–2 times per year (e.g., BIFAP, IPCI, LPD Italy). Nonetheless, all of these databases continue to accrue new information over time, adding data on new patients entering the healthcare system and updating data on existing patients who are followed longitudinally. There is substantial heterogeneity in how information is extracted for research across databases, and sometimes within databases, reflecting the diversity of software and healthcare systems. For example, in Spain, some autonomous regions send data to BIFAP directly, whereas other regions use local software programs for data extraction.

Primary care EHR databases generally contain a minimum set of clinical information, including data on patient demographics, medical diagnoses, and drug prescriptions. Table 13.2 contains detailed lists of data collected in each database and highlights differences among them.

European EHR databases use a variety of standardized coding systems to record diagnoses: Read codes (THIN and CPRD); International Classification of Diseases, 9th edition (ICD-9) (Arianna, one autonomous region in BIFAP, LPD Italy, and Pedianet); ICD-10 (IQVIA DA, SIDIAP, part of Pedianet, Hospital Episode Statistics [HES] data linked with CPRD and THIN); and the International Classification of Primary Care (ICPC) (IPCI, most autonomous regions in BIFAP).

The European EHR databases also vary in the ways they record drug data. CPRD and THIN employ British National Formulary (BNF) codes through the Gemscript system. European databases outside the UK record medications using Anatomical Therapeutic Chemical (ATC) classification codes, although many countries have national drug codes contained within a governmental formulary or similar compendium. All European EHR databases discussed in this chapter contain data on prescribed medications. Arianna, BIFAP, Pedianet, and SIDIAP also contain drug dispensing data. Additionally, Arianna, BIFAP, Pedianet, and (for roughly half of drugs) DA databases specify the indications for drugs.

All the above databases are representative of their respective source populations in terms of the distribution of age and sex and the prevalence of most diseases and prescribed drugs [18,19,22]. However, because data are collected for clinical and not research purposes, the reported frequency of certain diseases may vary across databases depending on local or disease-specific patterns of clinical care [38,39]. BIFAP, CPRD, DA France and Germany, IPCI, LPD Italy, and THIN include most regions of their respective countries. However, the distribution of patients across regions in these databases may not reflect the actual populations of those regions [1,15,17,40]. Similarly, the spectrum of socioeconomic status found in the databases may differ from the country as a whole [40].

Vital signs (e.g., blood pressure, height, weight, BMI) and laboratory test results are available in Arianna, BIFAP [41], CPRD [42], DA [43], IPCI, Pedianet, SIDIAP [44–46], and THIN [47] to varying degrees (see Table 13.2). For example, in both THIN and CPRD, laboratory data from after approximately 2000 is better recorded, but some older laboratory tests may not be available electronically if they were received by GPs on paper. Furthermore, height and weight are recorded for most adults in the UK but may be missing for many children (see Incompleteness of Clinical Data). In DA Germany, HbA1c for diabetic patients is nearly complete, but many other laboratory values are not recorded. Laboratory test results are also available in Arianna (~25% of patients), BIFAP, Pedianet (with indications for testing), and SIDIAP.

Hospitalizations, referrals, and the resulting consultation letters are recorded to varying degrees in European EHR databases. In the UK, discharge summaries and other hospital and consultant letters are sent to the GP, although these identified paper documents are not directly available to researchers. Linkage of THIN and CPRD to HES data allows researchers to access additional details from hospitalizations, such as diagnosis codes on admission and discharge and length of hospital stay. Referrals to other care settings are captured in both CPRD and THIN, and data from outside consultations may be obtained by linkage to HES data. Details on available data for hospitalizations and referrals in other databases are listed in Table 13.2.

Data from social history, including smoking and alcohol usage, are available to varying extents in most EHR databases (see Data Quality: Accuracy and Completeness, and Table 13.2). Substance exposure information is less consistently recorded in DA databases and IPCI. Pedianet contains information on parental smoking habits. Certain components of social history, such as occupation, are not routinely recorded in some databases [48].

In most of the European EHR databases described, most data are entered using structured (coded) fields rather than free text [49–52]. In contrast, BIFAP and IPCI contain large volumes of unstructured data. These and other databases make information from anonymized free text entries available to researchers (see Table 13.2). These free text data can be used to identify and validate outcomes and supplement available data in structured fields. Some practitioners may still keep paper record files, which could include precomputerization records, hospital discharge paperwork, or letters from specialists. Multiple databases, including THIN and CPRD, have additional fee-based data services that will obtain and anonymize paper-based data from GPs (see Table 13.2). To maximize the quality of the electronic record, IPCI does not permit participating GPs to keep paper records, but Dutch GPs record extensive free text notes, which are available to researchers [29].

Aside from the availability of unstructured data, several databases, including BIFAP, CPRD, DA Germany, Pedianet, SIDIAP, and THIN, allow researchers to administer questionnaires to clinicians or patients [51,53–55]. Like free text data, such questionnaires can be used to validate existing data or provide additional information that is not otherwise available in the database. Moreover, unlike free text entries, such questionnaires can be tailored in their content and administration based on the specific research question and population of interest [51,56]. Researchers must pay fees to administer supplementary questionnaires, a portion of which participants receive to complete the questionnaires. Of note, investigator-initiated surveys are also permitted in IPCI, but anecdotally response rates among GPs tend to be low.

Data Quality: Accuracy and Completeness

Data quality checks are performed by the European EHR databases at regular intervals on three levels: (1) practitioner recording, (2) data

extraction, and (3) maintenance of the database (see Table 13.1 for details on database-specific measures of quality assurance). When data are uploaded or extracted from the health records, the company processing the data performs additional quality checks to make sure the data have been correctly uploaded or extracted. Subsequent updates to the databases are verified for accuracy [1,40,57]. All databases undergo routine updating of the software used to collect, check, transfer, and present data.

In the UK, national quality improvement initiatives as well as advances in software have increased overall capture and accuracy of data. The national UK initiative the Quality and Outcomes Framework (QOF), a pay-for-performance program, was instituted in 2004 to improve performance using 146 quality indicators for 10 chronic diseases [58]. The QOF measures increased GP input in the EHR, leading to more complete data recording, especially for the targeted medical conditions [59,60]. Even after certain financial incentives were removed, performance for many quality indicators persisted across UK practices [61]. However, it is unclear whether reporting also improved for other, nonspecified quality indicators or diseases [62]. Some researchers have reported that the QOF has not contributed to decreases in mortality, better care coordination, or better patient experiences in the UK [63]. In its current form, therefore, this program's future is uncertain [64].

The QOF is only one of multiple quality improvement strategies implemented in the UK since the 1990s [62,65]. In coordination with specific databases, GPs receive training in the use of their software and regular evaluation of their data recording and prescribing behavior. GPs contributing to THIN or CPRD receive feedback reports with tips on improving recording and, in some cases, a summary of their prescribing habits relative to similar practices and other GPs in the UK. Other database-directed quality measures include audits of newly added practices and comparison of acquired data to national databases (e.g., mortality, hospitalizations, cancer, and cardiovascular registries) [22,66]. Finally, as incentives to take part in research studies, GPs and practices contributing to CPRD or THIN may receive income through questionnaires or participation in interventional clinical studies.

In CPRD, only data from practices that meet quality standards (~90% of practices) are provided for research. The Up-to-Standard (UTS) date is a practice-based quality marker corresponding to when a practice in CPRD is considered to have continuous and complete recording of data [17]. The UTS date is based on two parameters: the presence of gaps in the data stream and the existence of an appropriate rate of recorded deaths at the practice. For THIN, IQVIA employs a quality measure known as acceptable mortality reporting (AMR), denoting the year in which mortality reporting was deemed complete for each practice [40,53]. Other European EHR databases have their own standards for ensuring quality and completeness (see Table 13.1 – Quality checks, standards, and feedback for details).

With regard to specific variables, completeness of data varies among databases (see Incompleteness of Clinical Data and Table 13.2). Pregnancy, family structure, mortality, and cause of death are variably recorded and may be difficult to ascertain. Family structure may require the use of coding algorithms [66–74]. CPRD offers a probabilistic mother–baby link algorithm, which identifies likely mother–baby pairs based on an anonymized family number, maternity information from the mother's primary care record, and the month of birth of newly registered babies. CPRD may also be linked to mortality records from the Office for National Statistics to improve death estimates and confirm cause of death [75–77]. Researchers may also use algorithms for data in THIN to link family members or determine cause-specific mortality [78,79]. Risk factors such as smoking

and obesity may have gaps. The introduction of QOF measures led to a substantial increase in recording of these and other variables within CPRD and THIN [80–85]. For example, recording of data on adult patients' smoking habits rose from 75% before 2004 to nearly 90% by 2007 [86]. By 2005, 99% of patients with diabetes had a reported HbA1c value in the past 15 months compared to 87% in 1998.

As with other data sources, investigators need to consider the local context when interpreting dates in European EHR databases. Dates in the medical files may reflect dates of data entry or dates on which observations were made. In the case of new registrations in general practice, dates may reflect entry of data obtained from previous practitioners or from previous recording systems.

Data Access for Researchers

Research performed using European EHR databases must first be reviewed by the home institution's institutional review board (IRB) and the ethics board for the respective database. Given researchers' inability to identify individual patients in anonymized databases, such studies often meet the criteria for IRB exemption. Investigators must usually receive approval from the ethics board of the respective database before conducting their research. Requirements for approval change over time, so investigators should check with the data vendor about approval requirements prior to starting a study. Companies may also require completion of a data use agreement before initiation of a study. See Table 13.1 (Data access and approval) for more details.

United States: Department of Veterans Affairs Healthcare

Overview of Healthcare System and Population

The Department of Veterans Affairs (VA) was established in 1930 as the Veterans Administration based on congressional approval to "consolidate and coordinate Government activities affecting war veterans" [87,88]. The VA's Veterans Health Administration (VHA) is one of the largest integrated healthcare systems in the United States, providing medical, surgical, and rehabilitative care to a diverse group of military veterans as well as active duty reservists and National Guard. In contrast to the general US population, the VA population consists of predominantly older men (87% male, 47% over age 65 as of 2017) who often have multiple chronic medical or psychiatric conditions. The female population in the VA has increased over the last several years and represents a younger cohort of veterans. In 2016, the VA healthcare system consisted of 18 regional integrated networks encompassing 145 hospitals and medical centers, over 1200 ambulatory care, mobile, independent, and community-based outpatient clinics, and 132 community living centers (VA nursing homes) [89,90].

The VHA is primarily a direct provider of healthcare services, funded by the US government. While veterans receiving healthcare are not required to pay premiums for coverage, some are charged co-payments for certain medical services and outpatient prescriptions [88]. The vast majority of medications within the VHA are prescribed by VA clinicians and dispensed by VA pharmacies. To facilitate access to care, veterans may also see and receive medications from certain authorized private providers outside the VA [91]. Prescriptions from this program (<1% of all VA prescriptions as of 2017) are similarly dispensed and recorded within the VHA. Of note, dual-care veterans with Medicare coverage (which covers virtually all US citizens age 65 years and older) may receive medications through both the VA and the Medicare Part D Plan. This arrangement may result in duplicate or excessive drug usage for some patients – a clinically important situation that might be overlooked by clinicians and researchers alike who do not jointly consider both sources of medical care [91–93]. Because the VHA is not a

closed medical system, veterans may receive out-of-network care, compromising one's ability to study certain outcomes (see Incompleteness of Clinical Data).

In 2017, approximately 6.3 million veterans (more than one-third of US veterans) were treated in the VA healthcare system, with over 5 million receiving prescriptions (see Table 13.1 for details) [89].

Overview of Database

The VA database contains demographic, clinical, and administrative data from 1997 to the present along with prescription data since 1999 and laboratory data starting in 2002. This database contains health information on 14.5 million patients; in 2017, the number of veterans in the VA database was about 2% of the US population and 32% of US veterans [89].

Data Collection and Structure

Local VA medical centers record and store clinical and administrative data within the Veterans Health Information Systems Technology and Architecture (VistA) system. The National Patient Care Database (NPCD) contains inpatient data and (through 2018) outpatient data extracted, organized, and integrated from VistA. These data include demographics, diagnoses (ICD-9 and ICD-10 codes), clinic visits, admissions, discharges, transfers, prescription orders, laboratory, surgical procedures, provider specialty, and administrative services [94,95]. The VA's Corporate Data Warehouse (CDW), a primarily operational database, contains raw clinical, medication and administrative data through VistA [96,97]. While less structured than the NPCD, the CDW contains similar information as well as additional data, including vital signs, radiology results, free text notes, consults, and health factors (e.g., smoking status). Since 2019, the CDW has been the primary source of outpatient VA data.

These data may be used for quality improvement efforts and research. The VA Vital Status File contains some demographics (birth date, gender) and death data from multiple sources [98]. Death data are cross-checked monthly with the Social Security Administration Death Master File [98]. In addition, VA has several disease-specific registries that are used for patient care and research, including cancer, diabetes, severe mental illness, amyotrophic lateral sclerosis, rheumatoid arthritis, and the human immunodeficiency virus and hepatitis C clinical case registry.

Unlike most EHR databases, the VA database contains information on both prescribing and dispensing of drugs in both outpatient and inpatient settings. Pharmacy data systems record outpatient and inpatient drug dispensing in the CDW and Pharmacy Benefit Management (PBM) databases [99,100]. The PBM database also contains information on dispensing of nonprescription medications and specific medical supplies. The majority (85%) of outpatient medications are dispensed via the VA's consolidated mail-order pharmacies. While the PBM database records the dispensing of inpatient drugs, the Bar Code Medication Administration (BCMA) database contains records of administrations of medications to inpatients [101]. The VHA also maintains its own adverse drug event reporting system (also see Chapter 10); as of 2018, this system contains over 500 000 reports related to drugs or vaccines.

Investigators may also extract data for research directly from the EHR. Such primary data collection can facilitate, for example, the validation or ascertainment of outcomes in unstructured portions of the EHR, such as reports (e.g., biopsies), text-embedded test results (e.g., cardiac ejection fraction from an echocardiogram), and free text notes. Researchers may extract data through manual chart review in the local web-based EHR portal or via natural language processing programs [102,103]. In addition to EHR data, surveys of veterans and clinicians permit access to additional information [104].

Data Quality: Accuracy and Completeness

The CDW, a nontransformed mirror of the medical record, is updated nightly. Updates of the Vital Status File occur monthly. The accuracy and completeness of data reflect the EHR and beneficiary claims they source (see Tables 13.1 and 13.2 for more details). For instance, demographic race and ethnicity data can have up to 20% missing data in certain years [105].

The PBM database undergoes daily quality assurance processes to ensure completeness and accuracy [99]. As with medications obtained through Medicaid (see Chapter 12), low or nil co-payments produce strong financial incentives for veterans to obtain outpatient prescriptions through the VA.

Data Access for Researchers

Researchers may obtain data through any of the aforementioned VA data systems (see Data Collection and Structure) following approval by the local or central VA IRB. Access to VA data is limited to VA-affiliated researchers and their collaborators.

Strengths

Population-Based Data, Sample Size, and Length of Follow-up

Population-based studies draw subjects from the greater population to produce a derived sample that reflects the source [106]. Many European EHR databases allow researchers to use population-based study designs, minimizing selection bias and improving the validity and generalizability of epidemiologic studies. Although patients can opt out of having their information used, few do so.

Population-based data sources are ideal for nested case–control studies, in which all cases (e.g., individuals with the outcome) or a representative subsample are ascertained in a precisely defined population, and unaffected controls are sampled randomly from the same source population at the time when cases develop the outcome (incidence density sampling) [107]. Similarly, population-based data allow for the design of cohort studies, given the availability of prospective data with long follow-up periods. The large numbers of patients with longitudinal follow-up (see Table 13.1) may allow for sufficient statistical power to study rare exposures, diseases, and outcomes. These large, population-representative databases are excellent settings for a wide variety of methodologic and applied studies (see Particular Applications and Table 13.3).

Given the preponderance of older, sicker men in the VA system, the VA database is distinctly not representative of the US population or even of all US veterans, a majority of whom do not receive care within the VA system. However, the VA serves a high proportion of traditionally underrepresented and vulnerable groups, including the elderly and those with multiple co-morbidities, mental illness, disabilities, and lower socioeconomic status. The large and growing size of the VA population and retention of patients in the system, often until their death, facilitate large longitudinal studies within this special population of veterans.

Validity of Clinical Information

Epidemiologic studies in any of the EHR databases involve use of lists of codes, and sometimes algorithms, for specific medical conditions, drugs or other exposures of interest, and co-variates. Methods for deriving such code lists have been described [49,108]. The validity of such code lists and algorithms has been extensively studied in many of these databases. Studies of agreement between recording in the EHR and capture of data (e.g., prescription medications and specialist referrals) have been performed for certain databases [17,22,50,66,67,109]. Numerous studies have validated outcomes within EHR databases (see Table 13.3). For unvalidated outcomes of

Table 13.3 Examples of studies using EHR databases

Type of research	Setting or subject area	Sample of publications
Validation studies	BIFAP	Community-acquired pneumonia [189], ischemic stroke [190], meningioma [191], myocardial infarction [192], Stevens–Johnson syndrome and toxic epidermal necrolysis [193], and upper gastrointestinal bleeding [194]
	CPRD [17]	Atrial fibrillation [195], autism [196], cancer [197], cataract [198], chronic obstructive pulmonary disease [199–201], familial hypercholesterolemia [202], inflammatory bowel disease [203], liver injury [204], lymphoma [205], myocardial infarction and heart disease [206,207], pregnancy outcomes [68,69], pressure ulcers [208], psoriasis [48,209,210], psychosis [211], rheumatoid arthritis and juvenile idiopathic arthritis [212], Stevens–Johnson syndrome and toxic epidermal necrolysis [213], suicide [214], and venous thromboembolism [215]
	IQVIA DA databases [22]	Venous thromboembolism [216], general validation of pharmacoepidemiologic and pharmacoeconomic studies [22]
	LPD Italy	Heart failure, ischemic heart disease, hypertension, and type 2 diabetes [3]
	THIN [16]	Quality of cancer reporting [1,217], date of death and mortality reporting [53], hepatitis C virus infection [51], myocardial infarction [50], nonmelanomatous skin cancer [54], peptic ulcer disease [50,185], psoriasis [210], and stroke [50]
	VA	Acute kidney injury [218], fatty liver disease [219], heart failure [220], hepatocellular cancer [221], inflammatory bowel disease [222], myocardial infarction and related cardiac procedures [223], posttraumatic stress disorder [224], sepsis [225], and stroke [226]
Methodologic studies	Database-specific research (CPRD or THIN)	Timing and validity of diagnoses and outcomes relative to EHR-specific administrative dates [133,227,228] and clinical coding practices [229–231]; methods to impute missing drug information [134]; potential for misclassification and resultant bias due to missing data within free text [232,233], paper records, [234] and linked hospitalization records [235]
	Generalizable research using EHR databases (various)	Sources and types of bias [236–240]; novel study designs, such as self-controlled designs [241], regression discontinuity designs [242], and prevalent new-user cohort designs [243]; various analytic approaches for causal inference, such as propensity scores [244], high-dimensional propensity scores [245], simulation [246], marginal structural models [247,248], targeted maximum likelihood estimation [249], and instrumental variable analysis [250]; handling of missing data [251–253]; handling of repeated data [254,255]; and identification of outliers [256]

Table 13.3 (Continued)

Type of research	Setting or subject area	Sample of publications
Applied studies	General epidemiologic studies	Europe: representative incidence and prevalence studies [177,195,257–266] (e.g., shoulder complaints in primary care [267], newly diagnosed heart failure [268], bullous pemphigoid [269], and pemphigus vulgaris [269]); natural history of disease (e.g., irritable bowel syndrome [270]); risk of disease-related outcomes [260,271–279] (e.g., lymphoma among inflammatory bowel disease patients [205], myocardial infarction in patients with psoriasis [274], and complications of diabetes [280]); research on associated conditions [208,281] (e.g., obesity and liver disease [282]); patterns of diseases or symptoms [199]; rates of referral (e.g., chronic pelvic pain [283,284]); impact of geography [285,286] and pollution [287–289] on disease incidence and outcomes US VA: burden of illness associated with irritable bowel syndrome [290], military sexual trauma [291], patients awaiting major joint arthroplasty [292], and mental illness among veterans [293]
	Pharmacoepidemiologic studies	Europe: studies assessing risks [113,197,214,279,294–304] and outcomes [305–311] of medication (e.g., risk of myopathy and myalgias by statins [299]); safety and tolerability of medications [14,29,33,248,298,312–319]; studies of medication exposure and pregnancy outcomes [71,320,321]; reduction of morbidity or mortality by medication [306,322,323] and vaccinations [324]; persistence of medication use [325–327] (e.g., antihypertensives [328,329], bisphosphonates [330,331], and glaucoma therapies [332]); compliance and adherence [24,25,333,334]; physician's use of guidelines in prescribing medications [335–338] (e.g., antibiotics in children [339,340], antidepressants [341]); trends in prescribing [342–351]; device utilization [352,353], effectiveness [354], and safety [355,356] US VA: risks of myocardial infarction or musculoskeletal pain associated with bisphosphonates [357], dysglycemia with fluoroquinolones [358], neuropsychiatric adverse events with smoking cessation therapy [359], gastrointestinal bleeding with selective serotonin reuptake inhibitors [360], glucocorticoid-induced osteoporosis [361], and antipsychotic-associated mortality in dementia patients [362]
	Pharmacoeconomics, health services research, and pharmacovigilance	Europe: cost-effectiveness and safety of bisphosphonates [363], comparison of cost between glaucoma therapies [332,364], use and cost-effectiveness of long-term hormonal contraception [26,365], cost-effectiveness of treatment of gastroesophageal reflux disease [366]; health insurance-related barriers in access to and compliance with medicines [367–370]; healthcare utilization in fibromyalgia [371] and diabetes [30,372–376]; prescribing trends and their financial impact [377,378]; comparison of care of the elderly and nonelderly regarding symptoms concerning for ovarian cancer [379]; research on disparities and health outcomes [285,380–384]; variability in resource utilization and prescribing [385,386]; vaccination uptake and distribution [387–389]; impact of risk minimization measures [27,31] US VA: costs of erythropoietin therapy [390,391], treatment for metastatic prostate cancer [392], atrial fibrillation and stroke prevention [393], and healthcare costs of a collaborative intervention for chronic pain [394]; impact of clinical practice guidelines on quality of care [395] and adherence to the diabetes guidelines [396]

interest, one should strongly consider validating outcomes before or during a study to ensure that the specified diagnostic codes or algorithms reflect patients' true conditions. Codes or diagnoses that have not been validated may lead to spurious results and compromise a study's validity [50,51] (see also Chapter 37).

Accuracy of Drug Information

Electronic health record databases contain information on the name, strength, and quantity of prescribed drugs, which can be used to estimate their expected end date. In the UK, unlike in other countries, the prescription is the payment document. Although information is usually lacking on whether a prescribed drug has been dispensed from a UK pharmacy or taken by the patient, new prescriptions are generated and recorded when the current refills have been used and the patient requests a refill. A prespecified number of repeat prescriptions can be issued upon request of patients, after which a repeat clinical evaluation is required to ensure the prescribed therapy is still appropriate. The number of prespecified refills depends on factors such as patients' medical history and the drug in question.

Arianna, BIFAP, Pedianet, SIDIAP, and the VA databases contain information on outpatient drug dispensing in addition to the prescription data. VA databases also include detailed data on inpatient drug dispensing and administration. Of note, research using THIN has shown a high correspondence between issued and dispensed prescriptions except for a few selected drug classes; antipsychotics, drugs for malignancy, and immunosuppressants had lower redemption rates while anesthesia and vaccines were underreported as prescribed [110].

Ability to Access Original Health Records

Electronic health record databases contain information from patients' actual health records, which gives researchers insights into medical and social histories that are not possible with other types of databases (e.g., claims). For instance, researchers can access information about antecedent symptoms, prior medical conditions, family history, vital signs, physical exam finding, laboratory data, and prescribed medications, as detailed above. Particularly in closed medical systems where GPs and family pediatricians are the gatekeepers, health information in the EHR tends to be relatively comprehensive (see Europe and the United Kingdom, Overview of Healthcare Systems and Populations). Notably, some EHR databases permit access to anonymized free text data (e.g., IPCI), anonymized copies of paper records (e.g., THIN), or the entire EHR (e.g., VA), as well as access to clinicians or patients via surveys. These options allow researchers to verify information found elsewhere in a database or obtain additional supplemental data (see Data Collection and Structure). In published studies from the UK, response rates for health record requests have been greater than 80–90% [111–113].

Linkage to External Patient-Level Data

Many EHR databases may be linked to other health-related, patient-level information, thus extending the functionality and utility of EHR data. More than 75% of applications submitted to CPRD request the use of linked data (sometimes customized for a study) to augment the information available for research. The data source most commonly linked to CPRD and THIN is HES, which can provide data on hospital visits and stays, accident and emergency episodes, and tests by specialists, including imaging. The combination of data from primary care and HES facilitates research on conditions managed across multiple healthcare settings [114–117]. Linkage to official death records may improve the accuracy of mortality studies and validate data from general practice [118,119]. Researchers can link EHR data with other data sources, including disease (e.g.,

cancer) registries, mental health datasets, and socioeconomic and deprivation indices [120–124]. Furthermore, linkage of EHR data to individual patient-generated data is also possible, including patient-reported outcomes, environmental data, drug diaries, and biospecimens [125–127].

Electronic health record databases outside the UK also permit linkages to other data sources, including Arianna (e.g., claims data and registries on regional hospitalizations, national drug dispensing, and mortality), SIDIAP, and the VA (e.g., genetic information from the Million Veteran Program biobank [128] and administrative claims data from Medicare and Medicaid [92,129,130]). See Chapter 12 for information about other data sources that combine EHR and claims data.

Limitations

Incompleteness of Clinical Data

When using EHR databases for research, investigators rely on recording of patients' history and events by their clinicians and the health systems they work in. In clinical practice, human errors and omissions naturally occur, but systematic errors in recording can lead to bias. For example, in studies of the elderly, researchers can use available geriatric data to create a frailty score [131]. Notably, such data are likely to be selectively recorded for frailer patients [132]. Failure to account for such patterns of missingness could lead to bias. Although most EHR data are in electronic form, information received from outside sources (e.g., consultants or hospitals) in hard copy may not be fully captured if results are not manually reentered by clinicians. As with the above example, clinicians may be more likely to record laboratory or radiologic findings that are abnormal, but even abnormal results may not be documented reliably.

Because European EHR databases are designed to capture health information from primary care settings, they typically lack information from specialists. IQVIA DA databases record data from certain specialists, but these encounters are not linked directly to patients' primary care records. In other databases, recording of information from secondary care relies on communication between GPs and other clinicians. Researchers using THIN or CPRD may access more extensive and reliable data from other settings through linkage with HES data and other sources (see Linkage to External Patient-Level Data).

The nature of illness can affect the pattern of data recording. Unlike in administrative claims databases, codes for chronic diseases may be entered only once into EHR databases. For this reason, episodes of care involving acute events may be better recorded than chronic diseases [48,109,133]. Likewise, codes for chronic or inactive medical conditions may predate a study's follow-up or fixed baseline period. Failure to account for historical diagnoses could lead to misclassification of important study variables. Another consideration for studies using EHR data involves follow-up time and censoring: patients may transfer out of a given practice and entire practices may stop contributing patient-level data to the database. If such dropout relates both to exposure status and the outcome, bias could result.

While certain types of information can be found in EHR data and not administrative claims, some of these same variables often contain missing data. Examples include race and ethnicity, smoking and alcohol use, BMI, socioeconomic status, employment status, and occupation (see Tables 13.1 and 13.2 for database-specific details on percent recording). Unless incentives or other processes actively encourage recording (see Data Quality: Accuracy and Completeness), clinicians may more likely document these variables of interest (potential confounders) when they consider them relevant to patients' health. Of note, some databases, including CPRD, THIN, and SIDIAP, derive socioeconomic data from

patients' location of residence rather than their own finances.

Another source of missing data involves pediatric growth measurements in the UK. GPs routinely measure and record the height and weight of children in a paper record provided to families. However, GPs inconsistently document these same measurements electronically. As a result, longitudinal measures of pediatric growth are frequently incomplete in CPRD and THIN; as with other variables, recording of these measures may relate to other clinical characteristics. Recording of growth measurements in children is more comprehensive in Pedianet, although the overall size of Pedianet's pediatric population is smaller than that of either UK database.

Veterans in the VHA may receive healthcare outside the VHA either by choice (especially veterans aged 65 and older with Medicare coverage) or by necessity. For instance, veterans with urgent or emergent conditions are taken to the nearest hospital for care, as their VHA hospitals may be farther away or may lack a true emergency department. As a result, occurrences of acute conditions, such as myocardial infarction, stroke, and severe hypoglycemia, may be generally missed in inpatient data from VA hospitals, potentially resulting in missing outcome data. The frequent omission of acute outcomes of interest represents a major limitation of the VA database, making it less useful for some types of pharmacoepidemiologic studies when used as the only data source. Of note, in studies of veterans age 65 and older, one can overcome this limitation by linking VA data to Medicare claims data [91–93].

Incompleteness of Drug Data

Information on days' supply and daily dosage of prescribed medicines may not be explicitly recorded in EHR data. Methods are available for imputing days' supply [134]. Information obtained from the timing of repeat prescriptions or refills can inform the imputation of daily doses [135,136]. Additionally, algorithms have been developed to determine daily dosage and other drug data (e.g., frequency) from unstructured text [137–139]. Only a few databases (Arianna, BIFAP, Pedianet, and to some extent DA) specifically link prescribed drugs to a particular diagnosis. Without this information, one can refer to diagnoses recorded in or around encounters that correspond to prescribed drugs.

One must remember that prescribing records do not indicate which prescriptions were filled. Only a few of the EHR databases discussed (Arianna, BIFAP, Pedianet, SIDIAP, and the VA) also contain drug dispensing data, providing a more complete picture of drug utilization. Data on OTC drugs are frequently missing from EHR databases, but exceptions exist where healthcare systems pay for OTC drugs. For example, long-term use of certain OTC medications, such as aspirin and nonsteroidal antiinflammatory drugs, is recorded in UK databases [140] and Arianna [32,33]. The UK National Health Service provides free prescription items to certain segments of the population (e.g., patients over 60, children under 16, people in full-time education). Patients over age 60 also have free access to some chronically used nonprescription medications that are prescribed by GPs. These accommodations lead to more reliable usage data for OTC drugs in some UK populations. Additionally, Pedianet and SIDIAP capture prescribing and dispensing of drugs irrespective of coverage by the respective national health systems, leading to comprehensive recording of OTC medications. Notably, medication adherence is not well recorded in any setting; thus, documentation of prescribing or dispensing does not imply that patients actually take their medications (see Chapter 38 on Adherence).

In European EHR databases, data on medications restricted to specialist care,

dispensed from hospital pharmacies (e.g., biologics), and given during hospitalization or upon hospital discharge may be missing. In the UK, patients generally receive a limited quantity of medications upon hospital discharge, which in some cases may last two weeks. In the VA, certain medications are recorded in the EHR but are not accessible in the prescription databases, for example medications obtained from floor stock or administered acutely in acute care areas. The administration of other drugs may occasionally be incomplete due to work-arounds in documentation [141].

Complexity and Costs

The size and complexity of these databases require adequate computer hardware, software, and storage space as well as experienced data managers and analysts. These requirements carry costs in addition to the costs of using many databases themselves. Open-source software is available for managing and analyzing EHR data [142,143], including algorithms for use and analysis of free text [144,145].

Particular Applications

The aforementioned EHR databases have been a rich scientific setting for thousands of epidemiologic and pharmacoepidemiologic investigations. Research using these databases has included assessments of the incidence and natural history of disease; research on drug utilization, safety, and effectiveness; pharmacovigilance and signal detection; health economics research including cost-effectiveness; studies using case–control, cohort, longitudinal, self-controlled, and other designs; and a variety of methodologic studies. Examples of these studies are provided in Table 13.3. More comprehensive lists of publications using these databases may be found on their respective websites.

The Future

Electronic health records continue to evolve and expand in pursuit of delivering high-quality, high-value healthcare [146]. Interoperable EHR platforms and health information exchanges can enable broader access to and sharing of data across clinical settings. Such advances enable better coordination of care, reduce redundancy, and improve efficiency within and across healthcare systems. New technologies have improved communication not just among clinicians. Clinicians and their patients can now correspond electronically via the EHR and even interact in virtual clinical encounters through telemedicine. In some settings, patients can see their EHR data through patient-specific portals and upload their own data (e.g., patient-reported outcomes, consumer-wearable technology, "smart" digital technology) into the EHR. All these changes rely on and generate increasing volumes of data. Through advanced data analytics, healthcare systems can harness these data to learn from their patients, align and streamline processes of care, and improve patient outcomes – key objectives of learning health systems [147].

As EHR systems evolve, systems administrators, clinical informatics specialists, and end users must address important challenges. Missing data and variability in recording are common in EHRs. To optimize clinical care and maximize the potential for high-quality research, healthcare systems must implement approaches to ensure consistent and complete clinical documentation within the EHR. Future research will be needed to determine whether continuing education, financial incentive programs (e.g., QOF), targeted feedback, or other strategies can improve

patient outcomes while being cost-effective and sustainable.

In an age of ever-growing threats to data security, EHR databases require enhanced vigilance, technology, and standards to maintain individuals' privacy and confidentiality. Healthcare systems, industries, regulatory agencies, and governments must balance the potential societal benefits of broad access to and linking of health data with the personal risks to privacy and safety [148]. Contractual, practical, and technical barriers must be overcome to link and share data from disparate, proprietary EHR platforms.

The many advancements in EHR systems have important implications for the conduct of research. New technologic approaches, such as natural language processing, artificial intelligence, and other big-data analytics, have enabled researches to organize and use complex EHR data in novel ways [149–152]. Nonetheless, the clinical value of artificial intelligence-based methods to study medicines in observational settings remains unclear [153,154]. Large international research networks, such as ARITMO [29], SOS [155], and TEDDY [156] have demonstrated the capacity and power of international collaborations to use EHR databases for large-scale research on drug utilization and outcomes. Such collaborations lead to increased statistical power to study rare diseases, uncommonly used drugs, and rare outcomes; they also have greater external validity than research conducted within a single region or country. Linkage of EHRs with hospital data, administrative claims, and other data sources (e.g., patient registries) is increasingly important to maximize the advantages of each data source and minimize their respective limitations (see also Chapters 12 and 14) [92,116,157]. Furthermore, linkage of EHR data to patient-generated data and biospecimens adds to the possibilities for inquiry and discovery through population-representative, patient-centered research and molecular pharmacoepidemiology [128,158,159]. Researchers can also use the clinical networks and information infrastructure of EHR systems to conduct large pragmatic trials [160–162]. In addition, expansion of EHRs in low- and middle-income countries facilitates research and improvements in healthcare in areas of great need [163]. High-quality research conducted within EHR databases can favorably influence public policy and public health [164].

In the area of pharmacovigilance and regulation, EHR databases have an important role in postmarket evaluation. With rapid, nearly real-time analyses of recent population-based data, EHR databases are useful data sources to conduct postauthorization safety studies that monitor utilization and potential health risks of new drugs. Researchers, industries, and regulatory agencies may also use EHRs to quantify the impact of risk minimization measures, such as drug safety warnings issued by national drug agencies [165].

Through high-quality clinical and translational research and pharmacovigilance, EHR systems of the future will continue to serve as key platforms for answering important questions and improving the health of patients, communities, and populations.

Acknowledgment

The authors wish to thank the following individuals who provided information about specific databases: Arianna: Valentina Ientile; BIFAP: Miguel Gil García, Miguel Angel Maciá Martínez, Dolores Montero Corominas; HSD: Alessandro Pasqua, Iacopo Cricelli, Francesco Lapi; IPCI: Marcel de Wilde; Pedianet: Carlo Giaquinto, Luigi Cantarutti, Anna Cantarutti; SIDIAP: Bonaventura Bolíbar, Eduardo Hermosilla Pérez, Maria del Mar García Gil, Talita Duarte Salles; VA: Lucy Pandey, Kwan Hur. The authors also thank Molly E. Collins for her valuable editorial assistance.

References

1 Haynes K, Forde K, Schinnar R, Wong P, Strom B, Lewis J. Cancer incidence in The Health Improvement Network. *Pharmacoepidemiol Drug Saf* 2009; **18**: 730–6.

2 Peters SG, Buntrock JD. Big data and the electronic health record. *J Ambul Care Manage* 2014; **37**(3): 206–10.

3 Gini R, Schuemie MJ, Mazzaglia G, *et al*. Automatic identification of type 2 diabetes, hypertension, ischaemic heart disease, heart failure and their levels of severity from Italian General Practitioners' electronic medical records: a validation study. *BMJ Open* 2016; **6**(12): e012413.

4 Rivero-Calle I, Pardo-Seco J, Aldaz P, *et al*. Incidence and risk factor prevalence of community-acquired pneumonia in adults in primary care in Spain (NEUMO-ES-RISK project). *BMC Infect Dis* 2016; **16**(1): 645.

5 van Ginneken E. Perennial health care reform – the long Dutch quest for cost control and quality improvement. *N Engl J Med* 2015; **373**(10): 885–9.

6 García-Armesto S, Abadía-Taira MB, Durán A, Hernández-Quevedo C, Bernal-Delgado E. Health systems in transition: Spain, health system review. //www.euro.who.int/_data/assets/pdf_file/0004/128830/e94549.pdf (accessed April 9, 2019).

7 Heath D. France's health care system. *CMAJ* 2008; **178**(5): 596.

8 Lopes C. Health care in France: facing hard choices. *CMAJ* 2007; **177**(10): 1167–9.

9 Rodwin V. The health care system under French national health insurance: lessons for health reform in the United States. *Am J Public Health* 2003; **93**(1): 31–7.

10 Schlette S, Lisac M, Blum K. Integrated primary care in Germany: the road ahead. *Int J Integr Care* 2009; **9**: e14.

11 Walley T, Mantgani A. The UK General Practice Research Database. *Lancet* 1997; **350**: 1097–9.

12 Lawson D, Sherman V, Hollowell J. The General Practice Research Database. Scientific and Ethical Advisory Group. *Q J Med* 1998; **91**: 445–52.

13 Wood L, Coulson R. Revitalizing the General Practice Research Database: plans, challenges, and opportunities. *Pharmacoepidemiol Drug Saf* 2001; **10**: 379–83.

14 Wood L, Martinez C. The General Practice Research Database: role in pharmacovigilance. *Drug Saf* 2004; **27**(12): 871–81.

15 Wong I, Murray M. The potential of UK clinical databases in enhancing paediatric medication research. *Br J Clin Pharmacol* 2005; **59**(6): 750–5.

16 Lewis JD, Schinnar R, Bilker WB, Wang X, Strom BL. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2007; **16**(4): 393–401.

17 Herrett E, Thomas S, Schoonen S, Smeeth L, Hall A. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010; **69**(1): 4–14.

18 Herrett E, Gallagher AM, Bhaskaran K, *et al*. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44**(3): 827–36.

19 Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Inform Prim Care* 2011; **19**(4): 251–5.

20 Carbonari DM, Saine ME, Newcomb CW, *et al*. Use of demographic and pharmacy data to identify patients included within both the Clinical Practice Research Datalink (CPRD) and The Health Improvement Network (THIN). *Pharmacoepidemiol Drug Saf* 2015; **24**(9): 999–1003.

21 Cai B, Xu W, Bortnichak E, Watson DJ. An algorithm to identify medical practices common to both the General Practice Research Database and The Health Improvement Network database. *Pharmacoepidemiol Drug Saf* 2012; **21**(7): 770–4.

22 Becher H, Kostev K, Schroder-Bernhardi D. Validity and representativeness of the Disease Analyzer patient database for use in pharmacoepidemiological and pharmacoeconomic studies. *Int J Clin Pract* 2009; **47**(10): 617–26.

23 Dietlein G, Schroder-Bernhardi D. Use of the Mediplus patient database in healthcare research. *Int J Clin Pharmacol Ther* 2002; **40**(3): 130–3.

24 Guglielmi V, Bellia A, Pecchioli S, *et al*. Effectiveness of adherence to lipid lowering therapy on LDL-cholesterol in patients with very high cardiovascular risk: a real-world evidence study in primary care. *Atherosclerosis* 2017; **263**: 36–41.

25 Vetrano DL, Bianchini E, Onder G, *et al*. Poor adherence to chronic obstructive pulmonary disease medications in primary care: role of age, disease burden and polypharmacy. *Geriatrics Gerontol Int* 2017; **17**(12): 2500–6.

26 Lapi F, Simonetti M, Cricelli I, Cricelli C, Cassano N, Vena GA. Prescription appropriateness of cyproterone acetate/ethinylestradiol in primary care: a population-based study in Italy. *Clin Drug Invest* 2017; **37**(8): 755–62.

27 Sultana J, Fontana A, Giorgianni F, *et al*. The effect of safety warnings on antipsychotic drug prescribing in elderly persons with dementia in the United Kingdom and Italy: a population-based study. *CNS Drugs* 2016; **30**(11): 1097–109.

28 Tocci G, Nati G, Cricelli C, *et al*. Prevalence and control of hypertension in the general practice in Italy: updated analysis of a large database. *J Human Hypertens* 2017; **31**(4): 258–62.

29 Trifiro G, de Ridder M, Sultana J, *et al*. Use of azithromycin and risk of ventricular arrhythmia. *CMAJ* 2017; **189**(15): E560–E568.

30 Trifiro G, Parrino F, Pizzimenti V, *et al*. The management of diabetes mellitus in patients with chronic kidney disease: a population-based study in southern Italy. *Clin Drug Invest* 2016; **36**(3): 203–12.

31 Viola E, Trifiro G, Ingrasciotta Y, *et al*. Adverse drug reactions associated with off-label use of ketorolac, with particular focus on elderly patients. An analysis of the Italian pharmacovigilance database and a population based study. *Expert Opin Drug Saf* 2016; **15**(Suppl 2): 61–7.

32 Ingrasciotta Y, Sultana J, Giorgianni F, *et al*. The burden of nephrotoxic drug prescriptions in patients with chronic kidney disease: a retrospective population-based study in Southern Italy. *PloS One* 2014; **9**(2): e89072.

33 Ingrasciotta Y, Sultana J, Giorgianni F, *et al*. Association of individual non-steroidal anti-inflammatory drugs and chronic kidney disease: a population-based case control study. *PloS On.* 2015; **10**(4): e0122899.

34 Sultana J, Musazzi UM, Ingrasciotta Y, *et al*. Medication is an additional source of phosphate intake in chronic kidney disease patients. *Nutr Metab Cardiovasc Dis* 2015; **25**(10): 959–67.

35 Guerriero F, Orlando V, Tari DU, *et al*. How healthy is community-dwelling elderly population? Results from Southern Italy. *Transl Med UniSa* 2015; **13**: 59–64.

36 Brauer R, Ruigomez A, Downey G, *et al*. Prevalence of antibiotic use: a comparison across various European health care data sources. *Pharmacoepidemiol Drug Saf* 2016; **25**(Suppl 1): 11–20.

37 Ruigomez A, Brauer R, Rodriguez LA, *et al*. Ascertainment of acute liver injury in two European primary care databases. *Eur J Clin Pharmacol* 2014; **70**(10): 1227–35.

38 Jordan K, Clarke AM, Symmons DP, *et al*. Measuring disease prevalence: a comparison

of musculoskeletal disease using four general practice consultation databases. *Br J Gen Pract* 2007; **57**(534): 7–14.

39 Nightingale AL, Farmer RDT, de Vries CS. Systemic lupus erythematosus prevalence in the UK: methodological issues when using the General Practice Research Database to estimate frequency of chronic relapsing-remitting disease. *Pharmacoepidemiol Drug Saf* 2007; **16**(2): 144–51.

40 Maguire A, Blak B, Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiol Drug Saf* 2009; **18**: 76–83.

41 Martin-Merino E, Huerta-Alvarez C, Prieto-Alhambra D, Montero-Corominas D. Cessation rate of anti-osteoporosis treatments and risk factors in Spanish primary care settings: a population-based cohort analysis. *Arch Osteoporos* 2017; **12**(1): 39.

42 Cao X, Ejzykowicz F, Ramey DR, *et al*. Impact of switching from high-efficacy lipid-lowering therapies to generic simvastatin on LDL-C levels and LDL-C goal attainment among high-risk primary and secondary prevention populations in the United Kingdom. *Clin Therapeut* 2015; **37**(4): 804–15.

43 Kostev K, Parhofer KG, Dippel FW. Prevalence of high-risk cardiovascular patients with therapy-resistant hypercholesterolemia. *Cardiovasc Endocrinol* 2017; **6**(2): 81–5.

44 Ramos R, Ballo E, Marrugat J, *et al*. Validity for use in research on vascular diseases of the SIDIAP (Information System for the Development of Research in Primary Care): the EMMA study. *Rev Esp Cardiol* 2012; **65**(1): 29–37.

45 Vinagre I, Mata-Cases M, Hermosilla E, *et al*. Control of glycemia and cardiovascular risk factors in patients with type 2 diabetes in primary care in Catalonia (Spain). *Diabetes Care* 2012; **35**(4): 774–9.

46 Salvador-Gonzalez B, Mestre-Ferrer J, Soler-Vila M, *et al*. Chronic kidney disease in hypertensive subjects >/=60 years treated in primary care. *Nefrologia* 2017; **37**(4): 406–14.

47 Gunathilake W, Song S, Sridharan S, Fernando DJ, Idris I. Cardiovascular and metabolic risk profiles in young and old patients with type 2 diabetes. *QJM* 2010; **103**(11): 881–4.

48 Gelfand J, Margolis D, Dattani H. The UK General Practice Research Database. In: Strom B, ed. Pharmacoepidemiology, 4th edn. Chichester: John Wiley & Sons, 2005, pp. 337–46.

49 Dave S, Peterson I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf* 2009; **18**: 704–7.

50 Lewis J, Schinnar R, Bilker W, Wang X, Strom B. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiology Drug Saf* 2007; **16**: 393–401.

51 Lo Re V, Haynes K, Forde K, Localio A, Schinnar R, Lewis J. Validity of The Health Improvement Network (THIN) for epidemiologic studies of hepatitis C virus infection. *Pharmacoepidemiol Drug Saf* 2009; **18**: 807–14.

52 Khan N, Perera R, Harper S, Rose P. Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. *BMC Fam Pract* 2010; **5**(11): 1.

53 Hall G. Validation of death and suicide recording on the THIN UK primary care database. *Pharmacoepidemiol Drug Saf* 2009; **18**: 120–31.

54 Meal A, Leonardi-Bee J, Smith C, Hubbard R, Bath-Hextall F. Validation of THIN data for non-melanoma skin cancer. *Qual Primary Care* 2008; **16**: 49–52.

55 Khan N, Harrison S, Rose P. Validity of diagnostic coding within the General Practice Research Database: a systemic review. *Br J Gen Pract* 2010; **60**(572): e128–e136.

56 Scott FI, Mamtani R, Haynes K, Goldberg DS, Mahmoud NN, Lewis JD. Validation of a coding algorithm for intra-abdominal surgeries and adhesion-related complications in an electronic medical records database. *Pharmacoepidemiol Drug Saf* 2016; **25**(4): 405–12.

57 deLusignan S, Stephens P, Naeema N, Majeed A. Does feedback improve the quality of computerized medical records in primary care? *J Am Med Informatics Assoc* 2002; **9**(4): 395–403.

58 Doran T, Fullwood C, Gravelle H, *et al*. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006; **355**: 375–84.

59 Crawley D, Ng A, Mainous A, Majeed A, Millett C. Impact of pay for performance on quality of chronic disease management by social class group in England. *J R Soc Med* 2009; **102**: 103–7.

60 McDonald R, Roland M. Pay for performance in primary care in England and California: comparison of unintended consequences. *Ann Fam Med* 2009; **7**: 121–7.

61 Kontopantelis E, Springate D, Reeves D, Ashcroft DM, Valderas JM, Doran T. Withdrawing performance indicators: retrospective analysis of general practice performance under UK Quality and Outcomes Framework. *BMJ* 2014; **348**: g330.

62 Campbell S, Reeves D, Kontopantelis E, Middleton E, Siddald B, Roland M. Quality of primary care in England with the introduction of pay for performance. *N Engl J Med* 2007; **357**(2): 181–90.

63 Forbes LJ, Marchand C, Doran T, Peckham S. The role of the Quality and Outcomes Framework in the care of long-term conditions: a systematic review. *Br J Gen Pract* 2017; **67**(664): e775–e784.

64 Marshall M, Roland M. The future of the Quality and Outcomes Framework in England. *BMJ* 2017; **359**: j4681.

65 Health and Social Care Information Centre. National Dementia and Antipsychotic Prescribing Audit: National Summary Report. https://digital.nhs.uk/catalogue/PUB06624 (accessed April 9, 2019).

66 Jick SS, Kaye JA, Vasilakis-Scaramozza C, *et al*. Validity of the general practice research database. *Pharmacotherapy* 2003; **23**(5): 686–9.

67 Jick H, Jick S, Derby L. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *BMJ* 1991; **302**(6779): 766–8.

68 Charlton R, Cunnington M, de Vries C, Weil J. Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. *Drug Saf* 2008; **31**(1): 39–51.

69 Devine S, West S, Andrews E, Tennis P, Hammad T, Eaton S. The identification of pregnancies within the general research database. *Pharmacoepidemiol Drug Saf* 2010; **19**(1): 45–50.

70 Martín-Merino E, Ruigómez A, Johansson S, Wallander M, García-Rodriguez L. Study of a cohort of patients newly diagnosed with depression in general practice: prevalence, incidence, comorbidity, and treatment patterns. *Prim Care Compan J Clin Psychiatry* 2010; **12**(1).

71 Tata L, Lewis S, Mckeever T, *et al*. Congenital malformation in children born to women with asthma. *Pharmacoepidemiol Drug Saf* 2007; **16**(S271).

72 Hardy J, Holford T, Hall G, Bracken M. Strategies for identifying pregnancies in the automated medical records of General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2004; **13**: 749–59.

73 Hardy J, Leaderer B, Holford T, Hall G, Bracken M. Safety of medications prescribed before and during early pregnancy in a cohort of 81,975 mothers from the UK General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2006; **15**(8): 555–64.

74 McKeever T, Lewis S, Smith C, Collins J, Heatlie H, Frischer M. Siblings, multiple births, and the incidence of allergic disease: a birth cohort study using the West Midlands general practice research database. *Thorax* 2001; **56**(10): 758–62.

75 Joseph RM, Movahedi M, Dixon WG, Symmons DP. Smoking-related mortality in patients with early rheumatoid arthritis: a retrospective cohort study using the Clinical Practice Research Datalink. *Arthritis Care Res (Hoboken)* 2016; **68**(11): 1598–606.

76 Lane DA, Skjoth F, Lip GYH, Larsen TB, Kotecha D. Temporal trends in incidence, prevalence, and mortality of atrial fibrillation in primary care. *J Am Heart Assoc* 2017; **6**(5): ii.

77 Stewart D, Han L, Doran T, McCambridge J. Alcohol consumption and all-cause mortality: an analysis of general practice database records for patients with long-term conditions. *J Epidemiol Community Healt.* 2017; **71**(8): 729–35.

78 Scott FI, Horton DB, Mamtani R, *et al*. Administration of antibiotics to children before age 2 years increases risk for childhood obesity. *Gastroenterology* 2016; **151**(1): 120–9 e125.

79 Ogdie A, Maliha S, Shin D, *et al*. Cause-specific mortality in patients with psoriatic arthritis and rheumatoid arthritis. *Rheumatology (Oxford)* 2017; **56**(6): 907–11.

80 Lester H. The UK Quality and Outcomes Framework. *BMJ* 2008; **337**: a2095.

81 Sutton M, Elder R, Guthrie B, Watt G. Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. *Health Economics* 2010; **19**(1): 1–13.

82 Wise J. Framework for NHS annual "quality accounts" is launched. *BMJ* 2009; **339**: b3890.

83 Booth HP, Prevost AT, Gulliford MC. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. *Pharmacoepidemiol Drug Saf* 2013; **22**(12): 1357–61.

84 Bhaskaran K, Forbes HJ, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). *BMJ Open* 2013; **3**(9): e003389.

85 Marston L, Carpenter JR, Walters KR, *et al*. Smoker, ex-smoker or non-smoker? The validity of routinely recorded smoking status in UK primary care: a cross-sectional study. *BMJ Open* 2014; **4**(4): e004958.

86 Lewis J, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. *Pharmacoepidemiol Drug Saf* 2004; **13**: 437–41.

87 US Department of Veterans Affairs. VA History. www.va.gov/about_va/vahistory.asp (accessed April 9, 2019)

88 Panangala SV. *Veterans Medical Care: FY 2011 Appropriations*. CRS Report for Congress. Washington, DC: DIANE Publishing, 2010.

89 National Center for Veterans Analysis and Statistics. NCVAS Pocket Cards. www.va.gov/vetdata/pocketcard/index.asp (accessed April 9, 2019).

90 Thomas KS, Cote D, Makineni R, *et al*. Change in VA community living centers 2004–2011: shifting long-term care to the community. *J Aging Soc Policy* 2018; **30**(2): 93–108.

91 Gellad WF. The Veterans Choice Act and dual health system use. *J Gen Intern Med* 2016; **31**(2): 153–4.

92 Thorpe JM, Thorpe CT, Gellad WF, *et al*. Dual health care system use and high-risk prescribing in patients with dementia: a national cohort study. *Ann Intern Med* 2017; **166**(3): 157–63.

93 Gellad WF, Thorpe JM, Zhao X, *et al*. Impact of dual use of Department of Veterans Affairs and Medicare Part D drug benefits on potentially unsafe opioid use. *Am J Public Health* 2018; **108**(2): 248–55.

94 US Department of Veteran Affairs. National Patient Care Database (NPCD). catalog.data.gov/dataset/national–patient–care–database–npcd (access via log-in only).

95 Robb MA, Racoosin JA, Worrall C, Chapman S, Coster T, Cunningham FE. Active surveillance of postmarket medical product safety in the Federal Partners' Collaboration. *Med Care* 2012; **50**(11): 948–53.

96 US Department of Veterans Affairs. 172VA10P2: VHA Corporate Data Warehouse–VA (79 FR 4377). 2014.

97 US Department of Veterans Affairs. Corporate Data Warehouse (CDW). www.data.va.gov/dataset/corporate–data–warehouse–cdw (accessed April 9, 2019).

98 Tarlov E. Ascertaining veterans' vital status and dates of death: the VHA Vital Status File and other data sources for mortality ascertainment in VA. www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=857 (accessed April 9, 2019)

99 Aspinall SL, Sales MM, Good CB, *et al*. Pharmacy benefits management in the Veterans Health Administration revisited: a decade of advancements, 2004–2014. *J Manage Care Specialty Pharmacy* 2016; **22**(9): 1058–63.

100 Smith MW, Joseph GJ. Pharmacy data in the VA health care system. *Med Care Res Rev* 2003; **60**(3 Suppl): 92S–123S.

101 US Department of Veterans Affairs. *Bar Code Medication Administration (BCMA): Manager's User Manual*. Washington, DC: US Department of Veterans Affairs, 2004.

102 Patterson OV, Freiberg MS, Skanderson M, Fodeh SJ, Brandt CA, DuVall SL. Unlocking echocardiogram measurements for heart disease research through natural language processing. *BMC Cardiovasc Disord* 2017; **17**(1): 151.

103 Schroeck FR, Patterson OV, Alba PR, *et al*. Development of a natural language processing engine to generate bladder cancer pathology data for health services research. *Urology* 2017; **110**: 84–91.

104 Kazis LE. The Veterans SF-36 Health Status Questionnaire: development and application in the Veterans Health Administration. *Medical Outcomes Trust Monitor* 2000; **5**(1): 1–14.

105 Gebregziabher M, Zhao Y, Axon N, Gilbert G, Echols C, Egede L. Lessons learned in dealing with missing race data: an empirical investigation. *J Biomet Biostat* 2012; **3**(3): 1–5.

106 Szklo M. Population-based cohort studies. *Epidemiol Rev* 1998; **20**(1).

107 Rothman K, Greenland S. *Modern Epidemiology*. Philadelphia: Lippincott-Raven, 1988, pp. 93–115.

108 Dregan A, Grieve A, van Staa T, Gulliford MC, eCRT Research Team. Potential application of item–response theory to interpretation of medical codes in electronic patient records. *BMC Med Res Methodol* 2011; **11**: 168.

109 Hollowell J. The General Practice Research Database: quality of morbidity data. *Popul Trends* 1997; **87**: 36–40.

110 Prescribing and Primary Care Services of the NHS Information Centre. Prescribing Compliance: A Review of the Proportion of Prescriptions Dispensed. https://digital.nhs.uk/data-and-information/publications/statistical/prescribing-compliance/prescribing-compliance-a-review-of-the-proportion-of-prescriptions-dispensed (accessed April 10, 2019).

111 Arellano F, Arana A, Wentworth C, Fernandez-Vidaurre C, Schlienger R, Conde E. Lymphoma among patients with atopic dermatitis and/or treated with topical immunosuppressants in the United Kingdom. *J Allergy Clin Immunol* 2009; **123**(5): 1111–16.

112 Bourke A, Dattani H, Robinson M. Feasibility study and methodology to create a quality evaluated database of primary care data. *Informat Primary Care* 2004; **12**(3): 171–7.

113 Van Staa T, Abenhaim L, Cooper C, Zhang B, Leufkens H. The use of a large pharmacoepidemiological database to study exposure to oral corticosteroids and risk of fractures: validation of study population and results. *Pharmacoepidemiol Drug Saf* 2000; **9**(5): 359–66.

114 Crooks CJ, Card TR, West J. Defining upper gastrointestinal bleeding from linked primary and secondary care data and the effect on occurrence and 28 day mortality. *BMC Health Serv Res* 2012; **12**: 392.

115 Herrett E, Shah AD, Boggon R, *et al*. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ.* 2013; **346**: f2350.

116 Millett ER, Quint JK, de Stavola BL, Smeeth L, Thomas SL. Improved incidence estimates from linked vs. stand-alone electronic health records. *J Clin Epidemiol* 2016; **75**: 66–9.

117 Berni E, de Voogd H, Halcox JP, *et al*. Risk of cardiovascular events, arrhythmia and all–cause mortality associated with clarithromycin versus alternative antibiotics prescribed for respiratory tract infections: a retrospective cohort study. *BMJ Open* 2017; **7**(1): e013398.

118 Gallagher AM, Williams T, Leufkens HG, de Vries F. The impact of the choice of data source in record linkage studies estimating mortality in venous thromboembolism. *PLoS One* 2016; **11**(2): e0148349.

119 Morgan C, Webb RT, Carr MJ, *et al*. Incidence, clinical management, and mortality risk following self harm among children and adolescents: cohort study in primary care. *BMJ* 2017; **359**: j4351.

120 Iyen-Omofoman B, Hubbard RB, Smith CJ, *et al*. The distribution of lung cancer across sectors of society in the United Kingdom: a study using national primary care data. *BMC Public Health* 2011; **11**: 857.

121 Curtis EM, van der Velde R, Moon RJ, *et al*. Epidemiology of fractures in the United Kingdom 1988–2012: variation with age, sex, geography, ethnicity and socioeconomic status. *Bone* 2016; **87**: 19–26.

122 Bailey SE, Ukoumunne OC, Shephard EA, Hamilton W. Clinical relevance of thrombocytosis in primary care: a prospective cohort study of cancer incidence using English electronic medical records and cancer registry data. *Br J Gen Pract* 2017; **67**(659): e405–e413.

123 Morris M, Woods LM, Bhaskaran K, Rachet B. Do pre-diagnosis primary care consultation patterns explain deprivation-specific differences in net survival among women with breast cancer? An examination of individually-linked data from the UK West Midlands cancer registry, national screening programme and Clinical Practice Research Datalink. *BMC Cancer* 2017; **17**(1): 155.

124 Young GJ, Harrison S, Turner EL, *et al*. Prostate-specific antigen (PSA) testing of men in UK general practice: a 10-year longitudinal cohort study. *BMJ Open* 2017; **7**(10): e017729.

125 Moore E, Chatzidiakou L, Jones RL, *et al*. Linking e-health records, patient-reported symptoms and environmental exposure data to characterise and model COPD exacerbations: protocol for the COPE study. *BMJ Open* 2016; **6**(7): e011330.

126 Joseph RM, Soames J, Wright M, Sultana K, van Staa TP, Dixon WG. Supplementing electronic health records through sample collection and patient diaries: a study set within a primary care research database. *Pharmacoepidemiol Drug Saf* 2018; **27**(2): 239–42.

127 O'Meara H, Carr DF, Evely J, *et al*. Electronic health records for biological sample collection: feasibility study of statin-induced myopathy using the Clinical Practice Research Datalink. *Br J Clin Pharmacol* 2014; **77**(5): 831–8.

128 Gaziano JM, Concato J, Brophy M, *et al*. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* 2016; **70**: 214–23.

129 Abraham NS, Castillo DL, Hartman C. National mortality following upper gastrointestinal or cardiovascular events in older veterans with recent nonsteroidal anti–inflammatory drug use. *Aliment Pharmacol Ther* 2008; **28**(1): 97–106.

130 Wong ES, Rinne ST, Hebert PL, Cook MA, Liu CF. Hospital distance and readmissions among VA–Medicare dual-enrolled veterans. *J Rural Health* 2016; **32**(4): 377–86.

131 Clegg A, Bates C, Young J, *et al*. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing* 2018; **47**(2): 319.

132 Sultana J, Fontana A, Giorgianni F, *et al*. Can information on functional and cognitive status improve short-term mortality risk prediction among community-dwelling older people? A cohort study using a UK primary care database. *Clin Epidemiol* 2018; **10**: 31–9.

133 Lewis J, Bilker W, Weinstein R, Strom B. The relationship between time since registration and measured incidence rates in the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2005; **14**: 443–51.

134 Lum KJ, Newcomb CW, Roy JA, *et al*. Evaluation of methods to estimate missing days' supply within pharmacy data of the Clinical Practice Research Datalink (CPRD) and The Health Improvement Network (THIN). *Eur J Clin Pharmacol* 2017; **73**(1): 115–23.

135 Meid AD, Heider D, Adler JB, *et al*. Comparative evaluation of methods approximating drug prescription durations in claims data: modeling, simulation, and application to real data. *Pharmacoepidemiol Drug Saf* 2016; **25**(12): 1434–42.

136 Stovring H, Pottegard A, Hallas J. Determining prescription durations based on the parametric waiting time distribution. *Pharmacoepidemiol Drug Saf* 2016; **25**(12): 1451–9.

137 Shah AD, Martinez C. An algorithm to derive a numerical daily dose from unstructured text dosage instructions. *Pharmacoepidemiol Drug Saf* 2006; **15**(3): 161–6.

138 Karystianis G, Sheppard T, Dixon WG, Nenadic G. Modelling and extraction of variability in free-text medication prescriptions from an anonymised primary care electronic medical record research database. *BMC Med Inform Decis Mak* 2016; **16**: 18.

139 Lu CC, Leng J, Cannon GW, *et al*. The use of natural language processing on narrative medication schedules to compute average weekly dose. *Pharmacoepidemiol Drug Saf* 2016; **25**(12): 1414–24.

140 Yang YX, Hennessy S, Propert K, Hwang W, Sedarat A, Lewis J. Chronic proton pump inhibitor therapy and the risk of colorectal cancer. *Gastroenterology* 2007; **133**: 748–54.

141 Staggers N, Iribarren S, Guo JW, Weir C. Evaluation of a BCMA's electronic medication administration record. *Western J Nurs Res* 2015; **37**(7): 899–921.

142 Egbring M, Kullak-Ublick GA, Russmann S. Phynx: an open source software solution supporting data management and web-based patient-level data review for drug safety studies in the general practice research database and other health care databases. *Pharmacoepidemiol Drug Saf* 2010; **19**(1): 38–44.

143 Springate DA, Parisi R, Olier I, Reeves D, Kontopantelis E. rEHR: an R package for manipulating and analysing Electronic Health Record data. *PloS One* 2017; **12**(2): e0171784.

144 Wang Z, Shah AD, Tate AR, Denaxas S, Shawe-Taylor J, Hemingway H. Extracting

diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning. *PloS One* 2012; **7**(1): e30412.

145 Shah AD, Martinez C, Hemingway H. The freetext matching algorithm: a computer program to extract diagnoses and causes of death from unstructured text in electronic health records. *BMC Med Inform Decis Mak* 2012; **12**: 88.

146 Evans RS. Electronic health records: then, now, and in the future. *Yearb Med Inform* 2016; **Suppl 1**: S48–61.

147 Friedman C, Rubin J, Brown J, *et al*. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J Am Med Inform Assoc* 2015; **22**(1): 43–50.

148 Rumbold JM, Pierscionek B. The effect of the General Data Protection Regulation on medical research. *J Med Internet Res* 2017; **19**(2): e47.

149 Nelson SD, Lu CC, Teng CC, *et al*. The use of natural language processing of infusion notes to identify outpatient infusions. *Pharmacoepidemiol Drug Saf* 2015; **24**(1): 86–92.

150 Walker AM, Zhou X, Ananthakrishnan AN, *et al*. Computer-assisted expert case definition in electronic health records. *Int J Med Inform* 2016; **86**: 62–70.

151 Garvin JH, Kalsy M, Brandt C, *et al*. An evolving ecosystem for natural language processing in Department of Veterans Affairs. *J Med Syst* 2017; **41**(2): 32.

152 Trivedi G, Pham P, Chapman WW, Hwa R, Wiebe J, Hochheiser H. NLPReViz: an interactive tool for natural language processing on clinical text. *J Am Med Inform Assoc* 2018; **25**(1): 81–7.

153 Trifiro G, Sultana J, Bate A. From Big Data to smart data for pharmacovigilance: the role of healthcare databases and other emerging sources. *Drug Saf* 2018; **41**(2): 143–9.

154 Bate A, Reynolds RF, Caubel P. The hope, hype and reality of Big Data for pharmacovigilance. *Ther Adv Drug Saf* 2018; **9**(1): 5–11.

155 Arfe A, Scotti L, Varas-Lorenzo C, *et al*. Non-steroidal anti-inflammatory drugs and risk of heart failure in four European countries: nested case–control study. *BMJ* 2016; **354**: i4857.

156 Hsia Y, Neubert A, Sturkenboom MC, *et al*. Comparison of antiepileptic drug prescribing in children in three European countries. *Epilepsia* 2010; **51**(5): 789–96.

157 Chu TP, Shah A, Walker D, Coleman MP. Where are the opportunities for an earlier diagnosis of primary intracranial tumours in children and young adults? *Eur J Paediatr Neurol* 2017; **21**(2): 388–95.

158 Wing K, Douglas I, Bhaskaran K, *et al*. Development of predictive genetic tests for improving the safety of new medicines: the utilization of routinely collected electronic health records. *Drug Discov Today* 2014; **19**(4): 361–6.

159 Reade S, Spencer K, Sergeant JC, *et al*. Cloudy with a chance of pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. *JMIR Mhealth Uhealth* 2017; **5**(3): e37.

160 Van Staa TP, Dyson L, McCann G, *et al*. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess* 2014; **18**(43): 1–146.

161 Juszczyk D, Charlton J, McDermott L, *et al*. Electronically delivered, multicomponent intervention to reduce unnecessary antibiotic prescribing for respiratory infections in primary care: a cluster randomised trial using electronic health records – REDUCE Trial study original protocol. *BMJ Open* 2016; **6**(8): e010892.

162 Lederle FA, Cushman WC, Ferguson RE, Brophy MT, Fiore Md LD. Chlorthalidone versus hydrochlorothiazide: a new kind of Veterans Affairs cooperative study. *Ann Intern Med* 2016; **165**(9): 663–4.

163 Aminpour F, Sadoughi F, Ahamdi M. Utilization of open source electronic health record around the world: a systematic review. *J Res Med Sci* 2014; **19**(1): 57–64.

164 Oyinlola JO, Campbell J, Kousoulis AA. Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidances. *BMC Health Serv Res* 2016; **16**: 299.

165 Goedecke T, Morales DR, Pacurariu A, Kurz X. Measuring the impact of medicines regulatory interventions – systematic review and methodological considerations. *Br J Clin Pharmacol* 2018; **84**(3): 419–33.

166 Kaguelidou F, de Bie S, Verhamme K, *et al*. New quality indicators for paediatric antibiotic prescribing in primary care: a population based cohort study in the United Kingdom, Italy and the Netherlands from 1995–2010. *Arch Dis Child* 2016; **101**: e1–e1.

167 Mathur R, Bhaskaran K, Chaturvedi N, *et al*. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J Public Health* 2014; **36**(4): 684–92.

168 Guglielmi V, Bellia A, Bianchini E, *et al*. Drug interactions in users of tablet vs. oral liquid levothyroxine formulations: a real-world evidence study in primary care. *Endocrine* 2018; **59**(3): 585–92.

169 Gadroen K, Kemmeren JM, Bruijning-Verhagen PC, *et al*. Baseline incidence of intussusception in early childhood before rotavirus vaccine introduction, the Netherlands, January 2008 to December 2012. *Euro Surveillance* 2017; **22**(25).

170 Zamora A, Masana L, Comas-Cufi M, *et al*. Familial hypercholesterolemia in a European Mediterranean population – prevalence and clinical data from 2.5 million primary care patients. *J Clin Lipid* 2017; **11**(4): 1013–22.

171 US Department of Veteran Affairs. Pharmacy Benefits Management Services. www.pbm.va.gov/nationalformulary.asp (accessed April 10, 2019).

172 Sturkenboom MC, Dieleman JP, Picelli G, *et al*. Prevalence and treatment of hypertensive patients with multiple concomitant cardiovascular risk factors in The Netherlands and Italy. *J Human Hypertens* 2008; **22**(10): 704–13.

173 Garcia-Gil M, Hermosilla E, Prieto-Alhambra D, *et al*. Construction and validation of a scoring system for the selection of high-quality data in a Spanish population primary care database (SIDIAP). *Inform Prim Care* 2011; **19**(3): 135–45.

174 Valkhoff VE, Schade R, te Jong GW, *et al*. Population-based analysis of non-steroidal anti-inflammatory drug use among children in four European countries in the SOS project: what size of data platforms and which study designs do we need to assess safety issues? *BMC Pediatr* 2013; **13**: 192.

175 Rodriguez LA, Tolosa LB, Ruigomez A, Johansson S, Wallander MA. Rheumatoid arthritis in UK primary care: incidence and prior morbidity. *Scand J Rheumatol* 2009; **38**(3): 173–7.

176 Mulnier H, Seaman H, Raleigh V, Soedamah-Muthu S, Colhoun H, Lawrenson R. Mortality in people with type 2 daibetes in the UK. *Diabet Med* 2006; **23**: 516–21.

177 Neimann A, Shin D, Wang X, Margolis D, Troxel A, Gelfand J. Prevalence of cardiovascular risk factors in patients with psoriasis. *J Am Acad Dermatol* 2006; **55**: 829–35.

178 Delaney J, Daskalopoulou S, Brophy J, Steele R, Opatrny L, Suissa S. Lifestyle variables and the risk of myocardial infarction in the General Practice Research Database. *BMC Cardiovasc Disord* 2007; **7**(38): 1–8.

179 Cazzola M, Calzetta L, Lauro D, *et al*. Asthma and COPD in an Italian adult population: role of BMI considering the

smoking habit. *Respir Med* 2013; **107**(9): 1417–22.

180 Lapi F, Cassano N, Pegoraro V, *et al*. Epidemiology of chronic spontaneous urticaria: results from a nationwide, population-based study in Italy. *Br J Dermatol* 2016; **174**(5): 996–1004.

181 Afonso AS, Verhamme KM, Sturkenboom MC, Brusselle GG. COPD in the general population: prevalence, incidence and survival. *Respir Med* 2011; **105**(12): 1872–84.

182 Farmer R, Lawrenson R, Todd J, Williams T, MacRae K. Oral contraceptives and venous thromboembolic disease. Analysis of the UK General Practice Research Database and the UK MediPlus Database. *Human Reprod Update* 1999; **5**(6): 688–706.

183 Szathowski L, Lewis S, McNeill A, Coleman T. Is smoking status routinely recorded when patients register with a new GP? *Fam Pract* 2010; **27**(6): 673–5

184 Langley T, Szatkowski L, Gibson J, *et al*. Validation of the Health Improvement Network (THIN) primary care database for monitoring prescriptions for smoking cessation medications. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 586–90.

185 Cai S, Garcia Rodriguez L, Masso-Gonzalez E, Hernandez-Diaz S. Uncomplicated peptic ulcer in the UK: trends from 1997–2005. *Aliment Pharmacol Ther* 2009; **30**(10): 1039–48.

186 Baena-Diez JM, Garcia-Gil M, Comas-Cufi M, *et al*. Association between chronic immune-mediated inflammatory diseases and cardiovascular risk. *Heart* 2018;**104**(2): 119–26.

187 Cheung K, Aarts N, Noordam R, *et al*. Antidepressant use and the risk of suicide: a population-based cohort study. *J Affect Disord.* 2015; **174**: 479–84.

188 Gonzalez-Perez A, Garcia-Rodriguez L. Prostate cancer risk among men with diabetes mellitus. *Cancer Causes Control* 2005; **16**: 1055–8.

189 Saiz LC, Garjon J, Gorricho J, Erviti J, Gil-Garcia MJ, Martin-Merino E. Validation and incidence of community-acquired pneumonia in patients with type 2 diabetes in the BIFAP database. *Epidemiol Infect* 2017; **145**(14): 3056–64.

190 Garcia-Poza P, de Abajo FJ, Gil MJ, Chacon A, Bryant V, Garcia-Rodriguez LA. Risk of ischemic stroke associated with non-steroidal anti-inflammatory drugs and paracetamol: a population-based case–control study. *J Thromb Haemost* 2015; **13**(5): 708–18.

191 Gil M, Oliva B, Timoner J, Macia MA, Bryant V, de Abajo FJ. Risk of meningioma among users of high doses of cyproterone acetate as compared with the general population: evidence from a population-based cohort study. *Br J Clin Pharmacol* 2011; **72**(6): 965–8.

192 de Abajo FJ, Gil MJ, Garcia Poza P, *et al*. Risk of nonfatal acute myocardial infarction associated with non-steroidal antiinflammatory drugs, non-narcotic analgesics and other drugs used in osteoarthritis: a nested case–control study. *Pharmacoepidemiol Drug Saf* 2014; **23**(11): 1128–38.

193 Martin-Merino E, de Abajo FJ, Gil M. Risk of toxic epidermal necrolysis and Stevens–Johnson syndrome associated with benzodiazepines: a population-based cohort study. *Eur J Clin Pharmacol* 2015; **71**(6): 759–66.

194 de Abajo FJ, Gil MJ, Bryant V, Timoner J, Oliva B, Garcia-Rodriguez LA. Upper gastrointestinal bleeding associated with NSAIDs, other drugs and interactions: a nested case–control study in a new general practice database. *Eur J Clin Pharmacol* 2013; **69**(3): 691–701.

195 Ruigomez A, Johannson S, Wallander M, Rodriguez L. Incidence of chronic atrial fibrillation in general practice and its treatment pattern. *J Clin Epidemiol* 2002; **55**: 358–63.

196 Fombonne E, Heavey L, Smeeth L, Rodrigues L, Cook C, Smith P. Validation of the diagnosis of autism in general practitioner records. *BMC Public Health* 2004; **4**(5).

197 Jick H. Calcium-channel blockers and risk of cancer. *Lancet* 1997; **349**(9066): 1699–700.

198 Derby L, Maier W. Risk of cataract among users of intranasal corticosteroids. *J Allergy Clin Immunol* 2000; **105**(5): 912–16.

199 Soriano J, Maier W, Visick G, Pride N. Validation of general practitioner-diagnosed COPD in the UK General Practice Research Database. *Eur J Epidemiol* 2001; **17**: 1075–80.

200 Rothnie KJ, Mullerova H, Hurst JR, *et al*. Validation of the recording of acute exacerbations of COPD in UK primary care electronic healthcare records. *PLoS One* 2016; **11**(3): e0151357.

201 Rothnie KJ, Chandan JS, Goss HG, Mullerova H, Quint JK. Validity and interpretation of spirometric recordings to diagnose COPD in UK primary care. *Int J Chron Obstruct Pulmon Dis* 2017; **12**: 1663–8.

202 Weng SF, Kai J, Andrew Neil H, Humphries SE, Qureshi N. Improving identification of familial hypercholesterolaemia in primary care: derivation and validation of the familial hypercholesterolaemia case ascertainment tool (FAMCAT). *Atherosclerosis* 2015; **238**(2): 336–43.

203 Lewis J, Brensinger C, Bilker W, Strom B. Validity and completeness of the General Practice Research Database for studies of inflammatory bowel disease. *Pharmacoepidemiol Drug Saf* 2002; **11**: 211–18.

204 Wing K, Bhaskaran K, Smeeth L, *et al*. Optimising case detection within UK electronic health records: use of multiple linked databases for detecting liver injury. *BMJ Open* 2016; **6**(9): e012102.

205 Lewis J, Brensinger C, Bilker W, Deren J, Vaughn D, Strom B. Inflammatory bowel disease is not associated with an increased risk of lymphoma. *Gastroenterology* 2001; **121**: 1080–7.

206 Hammad TMM, Feight A, Iyasu S, Dal Pan G. Determining the predictive value of Read/OXMIS codes to identify incident acute myocardial infarction in the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2008; **17**(12): 1197–201.

207 Reeves D, Springate DA, Ashcroft DM, *et al*. Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with ischaemic heart disease: a cohort study with nested case–control analysis. *BMJ Open* 2014; **4**(4): e004952.

208 Margolis D, Bilker W, Knauss J, Baumgarten M, Strom B. The incidence and prevalence of pressure ulcers among elderly patients in general medical practice. *Ann Epidemiol* 2002; **12**(5): 321–5.

209 Gelfand J, Weinstein R, Porter S, Neimann A, Berlin J, Margolis D. Prevalence and treatment of psoriasis in the United Kingdom: a population-based study. *Arch Dermatol* 2005; **141**(12): 1537–41.

210 Seminara N, Abuabara K, Shin D, *et al*. Prevalence, treatment, and severity of psoriasis in The Health Improvement Network (THIN) – a population based study. *International Conference on Pharmacoepidemiology Conference,* 2010 (Abstract).

211 Nazareth I, King M, Haines A, Rangel L, Myers S. Accuracy of diagnosis of psychosis on general practice computer system. *BMJ* 1993; **307**(6895): 32–4.

212 Thomas S, Edwards C, Smeeth L, Cooper C, Hall A. How accurate are diagnoses for rheumatoid arthritis and juvenile idiopathic arthritis in the general practice research database? *Arthritis Rheum* 2008; **59**(9): 1314–21.

213 Frey N, Bircher A, Bodmer M, Jick SS, Meier CR, Spoendlin J. Validation of Stevens–

Johnson syndrome or toxic epidermal necrolysis diagnoses in the Clinical Practice Research Datalink. *Pharmacoepidemiol Drug Saf* 2017; **26**(4): 429–36.

214 Jick S, Dean A, Jick H. Antidepressants and suicide. *BMJ* 1995; **310**(6974): 215–18.

215 Lawrenson R, Todd J, Leydon G, Williams T, Farmer R. Validation of the diagnosis of venous thromboembolism in general practice database studies. *Pharmacoepidemiol Drug Saf* 2000; **49**(6): 591–6.

216 Todd J, Lawrenson R, Farmer R, Williams T, Leydon G. Venous thromboembolic disease and combined oral contraceptives: a re-analysis of the MediPlus database. *Human Reprod* 1999; **14**(6): 1500–5.

217 Arellano F, Conde E, Wenworth C, Schlienger R, Fernandez-Vidaurre C, Arana A. Validation of cases of lymphoma in THIN. *Pharmacoepi Drug Saf* 2008; **17**: S87–8.

218 Lafrance JP, Miller DR. Acute kidney injury associates with increased long-term mortality. *J Am Soc Nephrol* 2010; **21**(2): 345–52.

219 Husain N, Blais P, Kramer J, *et al*. Nonalcoholic fatty liver disease (NAFLD) in the Veterans Administration population: development and validation of an algorithm for NAFLD using automated data. *Aliment Pharmacol Ther* 2014; **40**(8): 949–54.

220 Mahajan SM, Burman P, Newton A, Heidenreich PA. A Validated risk model for 30-day readmission for heart failure. *Stud Health Technol Inform* 2017; **245**: 506–10.

221 Sada Y, Hou J, Richardson P, El-Serag H, Davila J. Validation of case finding algorithms for hepatocellular cancer from administrative data and electronic health records using natural language processing. *Med Care* 2016; **54**(2): e9–14.

222 Thirumurthi S, Chowdhury R, Richardson P, Abraham NS. Validation of ICD-9-CM diagnostic codes for inflammatory bowel disease among veterans. *Digest Dis Sci* 2010; **55**(9): 2592–8.

223 Petersen LA, Wright S, Normand SL, Daley J. Positive predictive value of the diagnosis of acute myocardial infarction in an administrative database. *J Gen Intern Med* 1999; **14**(9): 555–8.

224 Gravely AA, Cutting A, Nugent S, Grill J, Carlson K, Spoont M. Validity of PTSD diagnoses in VA administrative data: comparison of VA administrative PTSD diagnoses to self-reported PTSD Checklist scores. *J Rehab Res Dev* 2011; **48**(1): 21–30.

225 Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran's Affairs hospital discharge databases coded serious bacterial infections accurately. *J Clin Epidemiol* 2007; **60**(4): 397–409.

226 Niesner K, Murff HJ, Griffin MR, et al. Validation of VA administrative data algorithms for identifying cardiovascular disease hospitalization. *Epidemiology* 2013; **24**(2): 334–5.

227 Horsfall L, Walters K, Petersen I. Identifying periods of acceptable computer usage in primary care research databases. *Pharmacoepidemiol Drug Saf* 2013; **22**(1): 64–9.

228 Sammon CJ, Petersen I. Backdating of events in electronic primary health care data: should one censor at the date of last data collection. *Pharmacoepidemiol Drug Saf* 2016; **25**(4): 378–84.

229 Leite A, Andrews NJ, Thomas SL. Assessing recording delays in general practice records to inform near real-time vaccine safety surveillance using the Clinical Practice Research Datalink (CPRD). *Pharmacoepidemiol Drug Saf* 2017; **26**(4): 437–45.

230 Tate AR, Dungey S, Glew S, Beloff N, Williams R, Williams T. Quality of recording of diabetes in the UK: how does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study

using the CPRD database. *BMJ Open* 2017; **7**(1): e012905.

231 Tate AR, Martin AG, Murray-Thomas T, Anderson SR, Cassell JA. Determining the date of diagnosis – is it a simple matter? The impact of different approaches to dating diagnosis on estimates of delayed care for ovarian cancer in UK primary care. *BMC Med Res Methodo.* 2009; **9**: 42.

232 Ford E, Nicholson A, Koeling R, *et al*. Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is hidden in free text? *BMC Med Res Methodol* 2013; **13**: 105.

233 Price SJ, Stapley SA, Shephard E, Barraclough K, Hamilton WT. Is omission of free text records a possible source of data loss and bias in Clinical Practice Research Datalink studies? A case–control study. *BMJ Open* 2016; **6**(5): e011664.

234 Charlton RA, Weil JG, Cunnington MC, de Vries CS. Identifying major congenital malformations in the UK General Practice Research Database (GPRD): a study reporting on the sensitivity and added value of photocopied medical records and free text in the GPRD. *Drug Saf* 2010; **33**(9): 741–50.

235 Rothnie KJ, Mullerova H, Thomas SL, *et al*. Recording of hospitalizations for acute exacerbations of COPD in UK electronic health care records. *Clin Epidemiol* 2016; **8**: 771–82.

236 MacDonald TM, Morant SV, Goldstein JL, Burke TA, Pettitt D. Channelling bias and the incidence of gastrointestinal haemorrhage in users of meloxicam, coxibs, and older, non–specific non-steroidal anti-inflammatory drugs. *Gut* 2003; **52**(9): 1265–70.

237 Card TR, Solaymani-Dodaran M, Hubbard R, Logan RF, West J. Is an internal comparison better than using national data when estimating mortality in longitudinal studies? *J Epidemiol Community Health* 2006; **60**(9): 819–21.

238 Matok I, Azoulay L, Yin H, Suissa S. Immortal time bias in observational studies of drug effects in pregnancy. *Birth Defects Res A Clin Mol Teratol* 2014; **100**(9): 658–62.

239 Renoux C, dell'Aniello S, Brenner B, Suissa S. Bias from depletion of susceptibles: the example of hormone replacement therapy and the risk of venous thromboembolism. *Pharmacoepidemiol Drug Saf* 2017; **26**(5): 554–60.

240 Ankarfeldt MZ, Thorsted BL, Groenwold RH, Adalsteinsson E, Ali MS, Klungel OH. Assessment of channeling bias among initiators of glucose-lowering drugs: a UK cohort study. *Clin Epidemiol* 2017; **9**: 19–30.

241 Wang S, Linkletter C, Maclure M, *et al*. Future cases as present controls to adjust for exposure trend bias in case-only studies. *Epidemiology* 2011; **22**(4): 568–74.

242 Geneletti S, O'Keeffe AG, Sharples LD, Richardson S, Baio G. Bayesian regression discontinuity designs: incorporating clinical knowledge in the causal analysis of primary care data. *Stat Med* 2015; **34**(15): 2334–52.

243 Suissa S, Moodie EE, dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf* 2017; **26**(4): 459–68.

244 Morant SV, Pettitt D, MacDonald TM, Burke TA, Goldstein JL. Application of a propensity score to adjust for channelling bias with NSAIDs. *Pharmacoepidemiol Drug Saf* 2004; **13**(6): 345–53.

245 Toh S, Garcia Rodriguez LA, Hernan MA. Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records. *Pharmacoepidemiol Drug Saf* 2011; **20**(8): 849–57.

246 Tannen RL, Weiner MG, Xie D, Barnhart K. A simulation using data from a primary care practice database closely replicated the women's health initiative trial. *J Clin Epidemiol* 2007; **60**(7): 686–95.

247 Delaney JA, Daskalopoulou SS, Suissa S. Traditional versus marginal structural models to estimate the effectiveness of beta-blocker use on mortality after myocardial infarction. *Pharmacoepidemiol Drug Saf* 2009; **18**(1): 1–6.

248 Ali MS, Groenwold RH, Belitser SV, *et al*. Methodological comparison of marginal structural model, time-varying Cox regression, and propensity score methods: the example of antidepressant use and the risk of hip fracture. *Pharmacoepidemiol Drug Saf* 2016; **25**(Suppl 1): 114–21.

249 Pang M, Schuster T, Filion KB, Eberg M, Platt RW. Targeted maximum likelihood estimation for pharmacoepidemiologic research. *Epidemiology* 2016; **27**(4): 570–7.

250 Boef AG, Souverein PC, Vandenbroucke JP, *et al*. Instrumental variable analysis as a complementary analysis in studies of adverse effects: venous thromboembolism and second-generation versus third-generation oral contraceptives. *Pharmacoepidemiol Drug Saf* 2016; **25**(3): 317–24.

251 Delaney JA, Moodie EE, Suissa S. Validating the effects of drug treatment on blood pressure in the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2008; **17**(6): 535–45.

252 Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, Petersen I. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 618–26.

253 Welch CA, Petersen I, Bartlett JW, *et al*. Evaluation of two-fold fully conditional specification multiple imputation for longitudinal electronic health record data. *Stat Med* 2014; **33**(21): 3725–37.

254 Crowther MJ, Lambert PC, Abrams KR. Adjusting for measurement error in baseline prognostic biomarkers included in a time-to-event analysis: a joint modelling approach. *BMC Med Res Methodol* 2013; **13**: 146.

255 Sheppard T, Tamblyn R, Abrahamowicz M, Lunt M, Sperrin M, Dixon WG. A comparison of methods for estimating the temporal change in a continuous variable: example of HbA1c in patients with diabetes. *Pharmacoepidemiol Drug Saf* 2017; **26**(12): 1474–82.

256 Welch C, Petersen I, Walters K, *et al*. Two-stage method to remove population- and individual-level outliers from longitudinal data in a primary care database. *Pharmacoepidemiol Drug Saf* 2012; **21**(7): 725–32.

257 Linsell L, Dawson J, Zondervan K, *et al*. Prospective study of elderly people comparing treatments following first primary care consultation for a symptomatic hip or knee. *Fam Pract* 2005; **22**(1): 118–25.

258 Bath-Hextall F, Leonardi-Bee J, Smith C, Meal A, Hubbard R. Trends in incidence of skin basal cell carcinoma: additional evidence from a UK primary care database study. *Int J Cancer* 2007; **121**(9): 2105–8.

259 Van Staa T, Selby P, Leufkens H, Lyles K, Sprafka J, Cooper C. Incidence and natural history of Paget's disease of bone in England and Wales. *J Bone Min Res* 2002; **17**(3): 465–71.

260 Fischer L, Schlienger R, Matter C, Jick H, Meier C. Effect of rheumatoid arthritis or systemic lupus erythematosus on the risk of first-time acute myocardial infarction. *Am J Cardiol* 2004; **93**(2): 198–200.

261 Gupta D, Hansell A, Nichols T, Duong T, Ayres J, Strachan D. Epidemiology of pneumothorax in England. *Thorax* 2000; **55**(8): 666–71.

262 Humes D, Solaymani-Dodaran M, Fleming K, Simpson J, Spiller R, West J. A population-based study of perforated diverticular disease incidence and associated mortality. *Gastroenterology* 2009; **136**(4): 1198–205.

263 Punekar Y, Sheikh A. Establishing the incidence and prevalence of clinician-diagnosed allergic conditions in children and

adolescents using routinely collected data from general practices. *Clin Exp Allergy* 2009; **39**(8): 1209–16.

264 Souverein P, Webb D, Petri H, Weil J, Staa TV, Egberts T. Incidence of fractures among epilepsy patients: a population-based retrospective cohort study in the General Practice Research Database. *Epilepsia* 2005; **46**(2): 304–10.

265 Somers E, Thomas S, Smeeth L, Schoonen W, Hall A. Incidence of systemic lupus erythematosus in the United Kingdom; 1990–1999. *Arthritis Rheum* 2007; **57**(4): 612–18.

266 Davies A, Smeeth L, Grundy E. Contribution of changes in incidence and mortality to trends in the prevalence of coronary heart disease in the UK: 1996–2005. *Eur Heart J* 2007; **28**: 2142–7.

267 Linsell L, Dawson J, Zondervan K, *et al*. Prevalence and incidence of adults consulting for shoulder conditions in UK primary care; patters of diagnosis and referral. *Rheumatology* 2006; **45**: 215–21.

268 Johansson S, Wallander MA, Ruigomez A, Garcia Rodriguez LA. Incidence of newly diagnosed heart failure in UK general practice. *Eur J Heart Fail* 2001; **3**(2): 225–31.

269 Langan S, Smeeth L, Hubbard R, Fleming K, Smith C, West J. Bullous pemphigoid and pemphigus vulgaris – incidence and mortality in the UK: population based cohort study. *BMJ* 2008; **337**(a180).

270 Ruigomez A, Wallander MA, Johansson S, Garcia Rodriguez LA. One-year follow-up of newly diagnosed irritable bowel syndrome patients. *Aliment Pharmacol Ther* 1999; **13**(8): 1097–102.

271 Kaplan GG, Hubbard J, Korzenik J, *et al*. The inflammatory bowel diseases and ambient air pollution: a novel association. *Am J Gastroenterol* 2010; **105**(11): 2412–19.

272 Solaymani-Dodaran M, Aithal G, Card T, West J. Risk of cardiovascular and cerebrovascular events in primary biliary cirrhosis: a population-based cohort. *Am J Gastroenterol* 2008; **103**(11): 2784–8.

273 Clayton T, Thompson M, Meade T. Recent respiratory infection and risk of cardiovascular disease: case–control study through a general practice database. *Eur Heart J* 2008; **29**: 96–103.

274 Gelfand J, Neimann A, Shin D, Wang X, Margolis D, Troxel A. Risk of myocardial infarction in patients with psoriasis. *JAMA* 2006; **296**(14): 1735–42.

275 Pujades-Rodriguez M, Smith C, Hubbard R. Inhaled corticosteroids and the risk of fracture in chronic obstructive pulmonary disease. *Q J Med* 2007; **100**(8): 509–17.

276 Breart G, Cooper C, Meyer O, Speirs C, Deltour N, Reginster J. Osteoporosis and venous thromboembolism: a retrospective cohort study in the UK General Practice Research Database. *Osteoporos Int* 2010; **21**(7): 1181–7.

277 Hernán M, Jick S, Logroscino G, Olek M, Ascherio A, Jick H. Cigarette smoking and the progression of multiple sclerosis. *Brain* 2005; **128**(6): 1461–5.

278 Lewis N, Logan R, Hubbard R, West J. No increase in risk of fracture, malignancy or mortality in dermatitis herpetiformis: a cohort study. *Aliment Pharmacol Ther* 2008; **27**(11): 1140–7.

279 Van Staa T, Geusens P, Pols H, Laet CD, Leufkens H, Cooper C. A simple score for estimating the long-term risk of fracture in patients using oral glucocorticoids. *Q J Med* 2005; **98**: 191–8.

280 Huerta C, Zhao SZ, Garcia Rodriguez LA. Risk of acute liver injury in patients with diabetes. *Pharmacotherapy* 2002; **22**(9): 1091–6.

281 Davey G, Sedgwick P, Maier W, Visick G, Strachan D, Anderson H. Association between migraine and asthma: matched case–control study. *Br J Gen Pract* 2002; **52**(482): 723–7.

282 Meier C, Krahenbuhl S, Schlienger R, Jick H. Association between body mass index and liver disorders: an epidemiological study. *J Hepatol* 2002; **37**: 741–7.

283 Zondervan K, Yudkin P, Vessey M, Dawes M, Barlow D, Kennedy S. Patterns of diagnosis and referral in women consulting for chronic pelvic pain in UK primary care *Br J Obstet Gynaecol* 1999; **106**(11): 1156–61.

284 Zondervan K, Yudkin P, Vessey M, Dawes M, Barlow D, Kennedy S. Prevalence and incidence of chronic pelvic pain in primary care: evidence from a national general practice database. *Br J Obstet Gynaecol* 1999; **106**(11): 1149–55.

285 Perry DC, Bruce CE, Pope D, Dangerfield P, Platt MJ, Hall AJ. Legg–Calvé–Perthes disease in the UK: geographic and temporal trends in incidence reflecting differences in degree of deprivation in childhood. *Arthritis Rheum* 2012; **64**(5): 1673–9.

286 Shah A, Prieto-Alhambra D, Hawley S, *et al*. Geographic variation in secondary fracture prevention after a hip fracture during 1999–2013: a UK study. *Osteoporos Int* 2017; **28**(1): 169–78.

287 Hajat S, Haines A, Goubet SA, Atkinson RW, Anderson HR. Association of air pollution with daily GP consultations for asthma and other lower respiratory conditions in London. *Thorax* 1999; **54**(7): 597–605.

288 Atkinson RW, Carey IM, Kent AJ, van Staa TP, Anderson HR, Cook DG. Long-term exposure to outdoor air pollution and the incidence of chronic obstructive pulmonary disease in a national English cohort. *Occup Environ Med* 2015; **72**(1): 42–8.

289 Carey IM, Anderson HR, Atkinson RW, *et al*. Traffic pollution and the incidence of cardiorespiratory outcomes in an adult cohort in London. *Occup Environ Med* 2016; **73**(12): 849–56.

290 Spiegel BM. Burden of illness in irritable bowel syndrome: looking beyond the patient. *Clin Gastroenterol Hepatol* 2013; **11**(2): 156–7.

291 Haskell SG, Gordon KS, Mattocks K, *et al*. Gender differences in rates of depression, PTSD, pain, obesity, and military sexual trauma among Connecticut War Veterans of Iraq and Afghanistan. *J Womens Health* 2010; **19**(2): 267–71.

292 Kelly KD, Voaklander D, Kramer G, Johnston DW, Redfern L, Suarez-Almazor ME. The impact of health status on waiting time for major joint arthroplasty. *J Arthroplasty* 2000; **15**(7): 877–83.

293 Maynard C, Batten A, Liu CF, Nelson K, Fihn SD. The burden of mental illness among veterans: use of VHA health care services by those with service-connected conditions. *Med Care* 2017; **55**(11): 965–9.

294 Blumentals W, Brown R, Gomez-Caminero A. Antihypertensive treatment and erectile dysfunction in a cohort of type II diabetes patients. *Int J Impot Res* 2003; **15**(5): 314–17.

295 Kornegay C, Vasilakis-Scaramozza C, Jick H. Incident diabetes associated with antipsychotic use in the United Kingdom general practice research database. *J Clin Psychiatry* 2002; **63**(9): 758–62.

296 Van Staa T, Abenhaim L, Monette J. Rates of hypoglycemia in users of sulfonylureas. *J Clin Epidemiol* 1997; **50**(6): 735–41.

297 Langman M, Eichler E, Mavros P, Watson D, Kong S. Initiation of antihypertensive therapy among new users of cyclooxygenase-2-selective and nonselective NSAIDs. *Int J Clin Pharmacol Ther* 2004; **42**(5): 260–6.

298 Mochenhaupt M, Messenheimer J, Tennis P, Schlingmann J. Risk of Stevens–Johnson syndrome and toxic epidermal necrolysis in new users of antiepileptics. *Neurology* 2005; **64**: 1134–8.

299 Molokhia M, McKeigue P, Curcin V, Majeed A. Statin induced myopathy and myalgia: time trend analysis and comparison of risk associated with statin class from 1991–2006. *PLoS One* 2008; **3**(6): e2522.

300 Zhang H, Garcia-Rodriguez L, Hernandez-Diaz S. Antibiotic use and the risk of lung cancer. *Cancer Epidemiol* 2008; **17**(6): 1308–15.

301 Becker C, Jick S, Meier C. Use of statins and the risk of Parkinson's disease: a retrospective case–control study in the UK. *Drug Saf* 2008; **31**(5): 399–407.

302 Castellsague J, García-Rodríguez L, Duque A, Pérez S. Risk of serious skin disorders among users of oral antifungals: a population-based study. *BMC Dermatol* 2002; **2**: 14.

303 Meier C, Derby L, Jick S, Vasilakis C, Jick H. Antibiotics and risk of subsequent first-time acute myocardial infarction. *JAMA* 1999; **281**(5): 427–31.

304 Meier C, Wilcock K, Jick S. The risk of severe depression, psychosis or panic attacks with prophylactic antimalarials. *Drug Saf* 2004; **27**(3): 203–13.

305 Edwards C, Cooper C, Fisher D, Field M, Van Staa T, Arden N. The importance of the disease process and disease-modifying antirheumatic drug treatment in the development of septic arthritis in patients with rheumatoid arthritis. *Arthritis Rheum* 2007; **57**(7): 1151–7.

306 Bulugahapitiya U, Siyambalapitiya S, Sithole J, Fernando D, Idris I. Age threshold for vascular prophylaxis by aspirin in patients without diabetes. *Heart* 2008; **94**: 1429–32.

307 Gore M, Sadosky A, Leslie D, Sheehan A. Medication for treating neuropathic pain in patients with diabetes: a study using the UK and Germany Mediplus databases. *Pain Pract* 2008; **8**(4): 253–62.

308 Simons W, Vinod H, Gerber R, Bolinder B. Does rapid transition to insulin therapy in subjects with newly diagnosed type 2 diabetes mellitus benefit glycaemic control and diabetes-related complications? A German population-based study. *Exper Clin Endocrinol Diabetes* 2006; **114**(9): 520–6.

309 Smeeth L, Douglas I, Hall A, Hubbard R, Evans S. Effect of statins on a wide range of health outcomes: a cohort study validated by comparison with randomized trials. *Br J Clin Pharmacol* 2008; **67**(1): 99–109.

310 Thomas M, Williams A. Are outcomes the same with all dry powder inhalers? *Int J Clin Pract* 2005; **59**(Suppl 149): 33–5.

311 Zhang Q, Thomas M, Wisniewski T, Kocevar V, Price D. Treatment and outcomes in patients with asthma and allergic rhinitis in the UK. *Allergy Immunol.* 2007; **142**: 318–28.

312 McCarthy S, Cranswick N, Potts L, Taylor E, Wong I. Mortality associated with attention-deficit hyperactivity disorder (ADHD) drug treatment: a retrospective cohort study of children, adolescents and young adults using the general practice research database. *Drug Saf* 2009; **32**(11): 1089–96.

313 Hubbard R, Lewis S, West J, *et al*. Bupropion and the risk of sudden death: a self-controlled case-series analysis using The Health Improvement Network. *Thorax* 2005; **60**(10): 848–50.

314 Gaus W, Westendorf J, Diebow R, Kieser K. Identification of adverse drug reactions by evaluation of a prescription database, demonstrated for "risk of bleeding". *Methods Informat Med* 2005; **44**(5): 697–703.

315 Langman M, Kong S, Zhang Q, Kahler K, Finch E. Safety and patient tolerance of standard and slow-release formulations of NSAIDs. *Pharmacoepidemiol Drug Saf* 2003; **12**: 61–6.

316 Margolis D, Hoffstad O, Strom B. Association between serious ischemic cardiac outcomes and medications used to treat diabetes. *Pharmacoepidemiol Drug Saf* 2008; **17**: 753–9.

317 Tata L, Fortune P, Hubbard R, *et al*. Does concurrent prescription of selective serotonin reuptake inhibitors and non-steroidal anti-inflammatory drugs substantially increase the risk of upper

gastrointestinal bleeding? *Aliment Pharmacol Ther* 2005; **22**: 175–81.

318 de Bie S, Coloma PM, Ferrajolo C, *et al*. The role of electronic healthcare record databases in paediatric drug safety surveillance: a retrospective cohort study. *Br J Clin Pharmacol* 2015; **80**(2): 304–14.

319 Giner-Soriano M, Roso-Llorach A, Vedia Urgell C, *et al*. Effectiveness and safety of drugs used for stroke prevention in a cohort of non-valvular atrial fibrillation patients from a primary care electronic database. *Pharmacoepidemiol Drug Saf* 2017; **26**(1): 97–107.

320 Wood S, Jick H, Sauve R. The risk of stillbirth in pregnancies before and after the onset of diabetes. *Diabet Med* 2003; **20**: 703–7.

321 Zomerdijk IM, Ruiter R, Houweling LM, *et al*. Isotretinoin exposure during pregnancy: a population-based study in The Netherlands. *BMJ Open* 2014; **4**(11): e005602.

322 Garcia Rodriguez LA, Varas C, Patrono C. Differential effects of aspirin and non-aspirin nonsteroidal antiinflammatory drugs in the primary prevention of myocardial infarction in postmenopausal women. *Epidemiology* 2000; **11**(4): 382–7.

323 Gonzalez ELM, Rodriguez LAG. Proton pump inhibitors reduce the long-term risk of recurrent upper gastrointestinal bleeding: an observational study. *Aliment Pharmacol Therapeut* 2008; **28**(5): 629–37.

324 Schembri S, Winter J, MacDonald T. Influenza and pneumococcal vaccination protect against all cause mortality in patients with chronic obstructive pulmonary diseases. ISPE Abstract, Copenhagen, 2008.

325 Arellano FM, May C, Verma A, Cid J. Patterns of discontinuation, dose modification and switching among users of prescription non-steroidal anti-inflammatory drugs (NSAID) including cyclo-oxygenase 2 inhibitors (COX–2) in the UK population. *Pharmacoepidemiol Drug Saf* 2005; **14**(S1): 3.

326 Gopal M, Haynes K, Bellamy S, Arya L. Discontinuation rates of anticholinergic medications used for the treatment of lower urinary tract symptoms. *Obstet Gynecol* 2008; **112**(6): 1311–18.

327 Lyu R, Govoni M, Ding Q, *et al*. Treatment persistence among patients with rheumatoid disease (RA, AS, PsA) treated with subcutaneous biologics in Germany. *Rheumatol Int* 2016; **36**(1): 143–53.

328 Hasford J, Schroder-Bernhardi D, Rottenkolber M, Kostev K, Dietlein G. Persistence with antihypertensive treatments: results of a 3-year follow up cohort study. *Eur J Clin Pharmacol* 2007; **63**: 1055–61.

329 Suarez A, Staffa J, Fletcher P, Jones J. Reason for discontinuation of newly prescribed antihypertensive medications: methods of a pilot study using computerized patient records. *Pharmacoepidemiol Drug Saf* 2000; **9**: 405–16.

330 Cramer J, Gold D, Silverman S, Lewiecki E. A systematic review of persistence and compliance with bisphosphonates for osteoporosis. *Osteoporos Int* 2007; **18**(8): 1023–31.

331 Van Staa T, Geusens P, Leufkens H, Cooper C. Persistence to bisphosphonate treatment in actual clinical practice. *Bone* 2005; **36**(S2): S412.

332 Lafuma A, Berdeaux G. Costs and persistence of carbonic anhydrase inhibitor versus alpha-2 agonists, associated with beta-blockers, in glaucoma and ocular hypertension: an analysis of the UK GPRD database. *Curr Med Res Opin* 2008; **24**(5): 1519–27.

333 Schroder-Bernhardi D, Dietlein G. Compliance with prescription recommendations by physicians in practices. *Int J Clin Pharmacol Therapeut J* 2001; **39**(11): 477–9.

334 Ferrajolo C, Arcoraci V, Sullo MG, *et al*. Pattern of statin use in southern italian

primary care: can prescription databases be used for monitoring long-term adherence to the treatment? *PloS One* 2014; **9**(7): e102146.

335 Mikuls T, Farrar J, Bilker W, Fernandes S, Saag K. Suboptimal physician adherence to quality indicators for the management of gout and asymptomatic hyperuricaemia: results from the UK General Practice Research Database (GPRD). *Rheumatology* 2005; **44**(8): 1038–42.

336 Thompson PL, Gilbert RE, Long PF, Saxena S, Sharland M, Wong IC. Has UK guidance affected general practitioner antibiotic prescribing for otitis media in children? *J Public Health* 2008; **30**(4): 479–86.

337 Judge A, Wallace G, Prieto-Alhambra D, Arden NK, Edwards CJ. Can the publication of guidelines change the management of early rheumatoid arthritis? An interrupted time series analysis from the United Kingdom. *Rheumatology* 2015; **54**(12): 2244–8.

338 Kendrick T, Stuart B, Newell C, Geraghty AW, Moore M. Did NICE guidelines and the Quality Outcomes Framework change GP antidepressant prescribing in England? Observational study with time trend analyses 2003–2013. *J Affect Disord* 2015; **186**: 171–7.

339 Ferrajolo C, Verhamme KM, Trifiro G, *et al*. Antibiotic-induced liver injury in paediatric outpatients: a case–control study in primary care databases. *Drug Saf* 2017; **40**(4): 305–15.

340 Thompson P, Spyridis N, Sharland M, *et al*. Changes in clinical indications for community antibiotic prescribing for children in the UK from 1996 to 2006: will the new NICE prescribing guidance on upper respiratory tract infections just be ignored? *Arch Dis Child* 2009; **94**: 337–40.

341 Murray M, Thompson M, Santosh P, Wong I. Effects of the committee on safety of medicines advice on antidepressant prescribing to children and adolescents in the UK. *Drug Saf* 2005; **28**(12): 1151–7.

342 Munson J, Kreider M, Chen Z, Christie J, Kimmel S. Factors associated with the use of corticosteroids in the initial management of idiopathic pulmonary fibrosis. *Pharmacoepidemiol Drug Saf* 2010; **19**(7): 756–62.

343 Arellano F, Yood M, Wentworth C, *et al*. Use of cyclo-oxygenase 2 inhibitors (COX-2) and prescription non-steroidal anti-inflammatory drugs (NSAIDS) in UK and USA populations. Implications for COX-2 cardiovascular profile. *Pharmacoepidemiol Drug Saf* 2006; **15**: 861–72.

344 Boyc K, Yurgin N. Trends in the prescription of antidiabetic medications in France: evidence from primary care physicians. *Adv Ther* 2007; **24**(4): 803–14.

345 Sittle R, Nuijten M, Nautrup B. Patterns of dosage changes with transdermal buprenorphine and transdermal fentanyl for the treatment of noncancer and cancer pain: a retrospective data analysis in Germany. *Clin Therapeut* 2006; **28**(8): 1144–54.

346 Dietlein G, Schroder-Bernhardi D. Doctors' prescription behaviour regarding dosage recommendations for preparations of kava extracts. *Pharmacoepidemiol Drug Saf* 2003; **12**(5): 417–21.

347 Schroder-Bernhardi D, Dietlein G. Lipid-lowering therapy: do hospitals influence the prescribing behavior of general practitioners? *Int J Clin Pharmacol Therapeut* 2002; **40**(7): 317–21.

348 MacDonald T, Morant S. Prevalence and treatment of isolated and concurrent hypertension and hypercholesterolaemia in the United Kingdom. *Br J Clin Pharmacol* 2008; **65**(5): 775–86.

349 Turner S, Thomas M, Ziegenweidt Jv, Price D. Prescribing trends in asthma: a longitudinal observational study. *Arch Dis Child* 2009; **94**(1): 16–22.

350 Edwards C, Arden N, Fisher D, Saperia J, Reading I, Van Staa T. The changing use of disease-modifying anti-rheumatic drugs in

individuals with rheumatoid arthritis from the United Kingdom General Practice Research Database. *Rheumatology* 2005; **44**(11): 1394–8.

351 Suh D, Hunsche E, Shin H, Mavros P. Co-prescribing of proton pump inhibitors among chronic users of NSAIDs in the UK. *Rheumatology* 2008; **47**: 458–63.

352 Culliford DJ, Maskell J, Kiran A, *et al*. The lifetime risk of total hip and knee arthroplasty: results from the UK general practice research database. *Osteoarthritis Cartilage* 2012; **20**(6): 519–24.

353 Cea Soriano L, Wallander MA, Andersson S, Filonenko A, Garcia Rodriguez LA. Use of long-acting reversible contraceptives in the UK from 2004 to 2010: analysis using The Health Improvement Network Database. *Eur J Contracept Reprod Health Care* 2014; **19**(6): 439–47.

354 Guest JF, Fuller GW, Vowden P. Clinical outcomes and cost-effectiveness of three different compression systems in newly-diagnosed venous leg ulcers in the UK. *J Wound Care* 2017; **26**(5): 244–54.

355 Lalmohamed A, MacGregor AJ, de Vries F, Leufkens HG, van Staa TP. Patterns of risk of cancer in patients with metal-on-metal hip replacements versus other bearing surface types: a record linkage study between a prospective joint registry and general practice electronic health records in England. *PloS One* 2013; **8**(7): e65891.

356 Parker SE, Jick SS, Werler MM. Intrauterine device use and the risk of pre-eclampsia: a case–control study. *Br J Obstet Gynaecol* 2016; **123**(5): 788–95.

357 Caplan L, Pittman CB, Zeringue AL, *et al*. An observational study of musculoskeletal pain among patients receiving bisphosphonate therapy. *Mayo Clin Proc* 2010; **85**(4): 341–8.

358 Aspinall SL, Good CB, Jiang R, McCarren M, Dong D, Cunningham FE. Severe dysglycemia with the fluoroquinolones: a class effect? *Clin Infect Dis* 2009; **49**(3): 402–8.

359 Cunningham FE, Hur K, Dong D, *et al*. A comparison of neuropsychiatric adverse events during early treatment with varenicline or a nicotine patch. *Addiction* 2016; **111**(7): 1283–92.

360 Ayers K, Waljee AK, Saini SD. Short-term use of selective serotonin reuptake inhibitors increases risk of gastrointestinal bleeding in men with psychiatric diagnoses. *Gastroenterology* 2014; **147**(5): 1173–4.

361 Caplan L, Hines AE, Williams E, *et al*. An observational study of glucocorticoid-induced osteoporosis prophylaxis in a national cohort of male veterans with rheumatoid arthritis. *Osteoporos Int* 2011; **22**(1): 305–15.

362 Kales HC, Valenstein M, Kim HM, *et al*. Mortality risk in patients with dementia treated with antipsychotics versus other psychiatric medications. *Am J Psychiatry* 2007; **164**(10): 1568–76; quiz 1623.

363 Van Staa T, Geusens P, Leufkens H, Cooper C. An individualised model of cost-effectiveness of bisphosphonates in elderly women. *Bone* 2005; **36**(S2): S138.

364 Lafuma A, Berdeaux G. Costs and effectiveness of travoprost versus a dorzolamide + timolol fixed combination in first-line treatment of glaucoma: analysis conducted on the United Kingdom General Practitioner Research Database. *Curr Med Res Opin* 2007; **23**(12): 3009–16.

365 Varney S, Guest J. Relative cost effectiveness of Depo-Provera, Implanon, and Mirena in reversible long-term hormonal contraception in the UK. *Pharmacoeconomics* 2004; **22**(17): 1141–51.

366 Plumb J, Edwards S. Cost effectiveness analysis of proton pump inhibitors compared to omeprazole in the healing of reflux oesophagitis. *J Med Economics* 2002; **5**: 25–38.

367 Krobot K, Kaufman J, Christersen D, Preisser J, Miller W, Ibrahim M. Accessing a new

medication in Germany: a novel approach to assess a health insurance-related barrier. *Ann Epidemiol* 2005; **15**(10): 756–61.

368 Krobot K, Miller W, Kaufman J, Christ D. The disparity in access to new medication by type of health insurance: lessons from Germany. *Med Care* 2004; **42**(5): 487–91.

369 Krobot K, Miller W, Kaufman J, Christensen D, Preisser J, Ibrahim M. Quantifying delay in access to new medical treatment: an application of risk advancement period methodology. *Epidemiology* 2004; **15**(2): 202–7.

370 Atella V, Kopinska JA. The impact of cost-sharing schemes on drug compliance in Italy: evidence based on quantile regression. *Int J Public Health* 2014; **59**(2): 329–39.

371 Berger A, Sadosky A, Dukes E, Martin S, Edelsberg J, Oster G. Characteristics and patterns of healthcare utilization of patients with fibromyalgia in general practitioner settings in Germany. *Curr Med Res Opin* 2008; **24**(9): 2489–99.

372 Poole C, Tetlow T, McEwan P, Holmes P, Currie C. The prescription cost of managing people with type 1 and type 2 diabetes following initiation of treatment with either insulin glargine or insulin detemir in routine general practice in the UK: a retrospective database analysis. *Curr Med Res Opin* 2007; **8**: 41–8.

373 Rathmann W, Haastert B, Giani G. Drug prescriptions and costs in the treatment of diabetic polyneuropathy. *Dtsch Med Wschr* 1999; **124**: 681–6.

374 Icks A, Haastert B, Giani G, Rathmann W. Incremental prescription and drug costs during the years preceding diabetes diagnosis in primary care practices in Germany. *Exp Clin Endocrinol Diabetes* 2006; **114**: 348–55.

375 Gulliford M, Latinovic R, Charlton J. Diabetes diagnosis, resource utilization, and health outcomes. *Am J Manag Care* 2008; **14**(1): 32–8.

376 Mazzaglia G, Yurgin N, Boye KS, *et al*. Prevalence and antihyperglycemic prescribing trends for patients with type 2 diabetes in Italy: a 4-year retrospective study from national primary care data. *Pharmacol Res* 2008; **57**(5): 358–63.

377 Blak B, Mullins C, Shaya F, Simoni-Wastila L, Cooke C, Weir M. Prescribing trends and drug budget impact of the ARBs in the UK. *Value Health* 2009; **12**(2): 302–8.

378 Hughes D, McGuire A. The direct costs to the NHS of discontinuing and switching prescriptions for hypertension. *J Human Hypertens* 1998; **12**(8): 533–7.

379 Tate A, Nicholson A, Cassell J. Are GPs under-investigating older patients presenting with symptoms of ovarian cancer? Observational study using General Practice Research Database. *Br J Cancer* 2010; **102**(6): 947–51.

380 Keane MG, Horsfall LJ, Rait G, Pereira SP. Sociodemographic trends in the incidence of pancreatic and biliary tract cancer in UK primary care. *PloS One* 2014; **9**(9): e108498.

381 Garcia-Gil M, Elorza JM, Banque M, *et al*. Linking of primary care records to census data to study the association between socioeconomic status and cancer incidence in Southern Europe: a nation-wide ecological study. *PloS One* 2014; **9**(10): e109706.

382 Hamer HM, Kostev K. Sociodemographic disparities in administration of antiepileptic drugs to adults with epilepsy in Germany: a retrospective, database study of drug prescriptions. *CNS Drugs* 2014; **28**(8): 753–9.

383 Conrad N, Judge A, Tran J, *et al*. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *Lancet* 2018; **391**(10120): 572–80.

384 Klop C, van Staa TP, Cooper C, Harvey NC, de Vries F. The epidemiology of mortality after fracture in England: variation by age,

sex, time, geographic location, and ethnicity. *Osteoporos Int* 2017; **28**(1): 161–8.

385 Busby J, Schroeder K, Woltersdorf W, *et al*. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: using routine data to identify research priorities. *Br J Gen Pract* 2013; **63**(609): e256–66.

386 van der Velde RY, Wyers CE, Teesselink E, *et al*. Trends in oral anti-osteoporosis drug prescription in the United Kingdom between 1990 and 2012: variation by age, sex, geographic location and ethnicity. *Bone* 2017; **94**: 50–5.

387 Vermeer-de Bondt PE, Schoffelen T, Vanrolleghem AM, *et al*. Coverage of the 2011 Q fever vaccination campaign in the Netherlands, using retrospective population-based prevalence estimation of cardiovascular risk-conditions for chronic Q fever. *PloS One* 2015; **10**(4): e0123570.

388 Voordouw AC, Sturkenboom MC, Dieleman JP, *et al*. Annual revaccination against influenza and mortality risk in community-dwelling elderly persons. *JAMA* 2004; **292**(17): 2089–95.

389 Joseph C, Goddard N, Gelb D. Influenza vaccine uptake and distribution in England and Wales using data from the General Practice Research Database, 1989/90–2003/04. *J Public Health* 2005; **27**(4): 371–7.

390 Aspinall SL, Smith KJ, Good CB, *et al*. Incremental cost effectiveness of pharmacist-managed erythropoiesis-stimulating agent clinics for non-dialysis-dependent chronic kidney disease patients. *Appl Health Econ Health Policy* 2013; **11**(6): 653–60.

391 Hynes DM, Stroupe KT, Greer JW, *et al*. Potential cost savings of erythropoietin administration in end-stage renal disease. *Am J Med* 2002; **112**(3): 169–75.

392 Zeliadt SB, Penson DF. Pharmacoeconomics of available treatment options for metastatic prostate cancer. *Pharmacoeconomics* 2007; **25**(4): 309–27.

393 Lich KH, Tian Y, Beadles CA, *et al*. Strategic planning to reduce the burden of stroke among veterans: using simulation modeling to inform decision making. *Stroke* 2014; **45**(7): 2078–84.

394 Dickinson KC, Sharma R, Duckart JP, Corson K, Gerrity MS, Dobscha SK. VA healthcare costs of a collaborative intervention for chronic pain in primary care. *Med Care* 2010; **48**(1): 38–44.

395 Lin GA, Redberg RF, Anderson HV, *et al*. Impact of changes in clinical practice guidelines on assessment of quality of care. *Med Care* 2010; **48**(8): 733–8.

396 Conlin PR, Colburn J, Aron D, Pries RM, Tschanz MP, Pogach L. Synopsis of the 2017 U.S. Department of Veterans Affairs/U.S. Department of Defense Clinical Practice Guideline: Management of Type 2 Diabetes Mellitus. *Ann Intern Med* 2017; **167**(9): 655–63.