

RESEARCH ARTICLE

Open Access



Factors influencing the development of primary care data collection projects from electronic health records: a systematic review of the literature

Marie-Line Gentil^{1,2*} , Marc Cuggia^{3,4,5,6}, Laure Fiquet^{1,2}, Camille Hagenbourger¹, Thomas Le Berre¹, Agnès Banâtre^{1,2}, Eric Renault⁴, Guillaume Bouzille^{3,4,5,6} and Anthony Chapron^{1,2}

Abstract

Background: Primary care data gathered from Electronic Health Records are of the utmost interest considering the essential role of general practitioners (GPs) as coordinators of patient care. These data represent the synthesis of the patient history and also give a comprehensive picture of the population health status. Nevertheless, discrepancies between countries exist concerning routine data collection projects. Therefore, we wanted to identify elements that influence the development and durability of such projects.

Methods: A systematic review was conducted using the PubMed database to identify worldwide current primary care data collection projects. The gray literature was also searched via official project websites and their contact person was emailed to obtain information on the project managers. Data were retrieved from the included studies using a standardized form, screening four aspects: projects features, technological infrastructure, GPs' roles, data collection network organization.

Results: The literature search allowed identifying 36 routine data collection networks, mostly in English-speaking countries: CPRD and THIN in the United Kingdom, the Veterans Health Administration project in the United States, EMERALD and CPCSSN in Canada. These projects had in common the use of technical facilities that range from extraction tools to comprehensive computing platforms. Moreover, GPs initiated the extraction process and benefited from incentives for their participation. Finally, analysis of the literature data highlighted that governmental services, academic institutions, including departments of general practice, and software companies, are pivotal for the promotion and durability of primary care data collection projects.

Conclusion: Solid technical facilities and strong academic and governmental support are required for promoting and supporting long-term and wide-range primary care data collection projects.

Keywords: Primary care, Data mining, Data collection, Secondary use, Electronic health records, Governance, Stakeholders

* Correspondence: marieline.gentil@gmail.com

¹Department of General Practice, University of Rennes 1, F-35000 Rennes, France

²CIC (Clinical investigation center) INSERM 1414, F-35000 Rennes, France

Full list of author information is available at the end of the article



Background

The secondary use of Electronic Health Record (EHR) data, for instance for epidemiological research, pharmacovigilance or health policy making, is progressively increasing [1]. Moreover, due to the chronic nature of many diseases, a global understanding of the patient's history is crucial for quality healthcare. In this respect, a paradigm shift occurred with the development of Big Data analysis following EHR digitization that facilitates data processing. Indeed, data mining brings large amount of information with higher granularity (i.e., higher level of detail).

In France, several data retrieval projects [2–4] are currently focused on the collection and mining of hospital administrative data (for instance, the Program of Medicalization of the Information Systems) and of clinical data from hospital EHRs. Data retrieved from hospital sources are promising, but they do not take into account the entire care pathway of each single patient or of the whole population. From this point of view, primary care records are particularly interesting [5]. Indeed, as general practitioners (GPs) are often the coordinators for their patients' healthcare trajectory, primary care records should contain the entire medical history of each patient [6]. Moreover, most people have access to primary care. Thanks to information technologies (IT), the volume of data captured by EHRs, paired with the growing capacity for data linkage and exchange, creates opportunities for measuring outcomes and, consequently, for improving patient and population health. In France, one such initiative was the "Observatoire de la Médecine Générale" (a nationwide survey of GPs' practice) that ended in 2009 due to lack of funding. Currently, the French Institute for Research and Documentation on Health Economics (IRDES) exploits the primary care data obtained by other companies, such as IMS Health®, a private-sector firm [7], via partnership agreements. Nevertheless, the lack of information on how these data were collected raises methodological concerns [7]. Cegedim®, another private-sector company, works on data extracted from French primary care EHRs [8]. Few local initiatives also have been implemented [9], but we could not find a transparent French national infrastructure that collects data directly from primary care practices. Conversely, in other countries, the possible contribution to medical science and health policy decision-making of routinely collected primary care data is now assessed, for instance with the Clinical Practice Research Datalink (CPRD) in the United Kingdom (UK) [10, 11].

Considering the discrepancies between countries on routine data collection from primary care EHRs, we wanted to identify factors that might facilitate the development and durability of routine primary care data collection. To this aim, we reviewed primary care data

collection projects worldwide by taking into account their technical features, the GPs' contribution and the network managers.

Methods

A systematic review of the literature was performed from December 2015 to November 2016, based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA [12]) criteria.

The checklist points for Assessing the Methodological Quality of Systematic Reviews (AMSTAR [13]) were completed when criteria were relevant, that is to say points 1 to 5, 10 and 11. We referred to it at the beginning to define our work protocol (point 1), duplicate study selection and data extraction (point 2), perform a comprehensive literature search (point 3 and 5). Then it guided us through our analysis of included papers: status of publication (point 4), scientific quality, publication bias (point 10) and conflict of interest (point 11).

Information sources

First, an automated literature search of the PubMed database was performed with the assistance of a university librarian, with expertise in systematic reviews.

To identify data collection projects based on primary care EHRs, worldwide, our query was divided in three parts: i) collection of EHR synonyms; ii) retrieval of records about automatic data processing; and iii) identification of primary care data collection projects, using several MeSH term synonyms. To expand our search, MeSH terms and also free text words were used. Articles published from 2010 onwards were selected. No other filter was applied. The following query (Fig. 1) was submitted to the PubMed search engine. The last search was performed in November 2016.

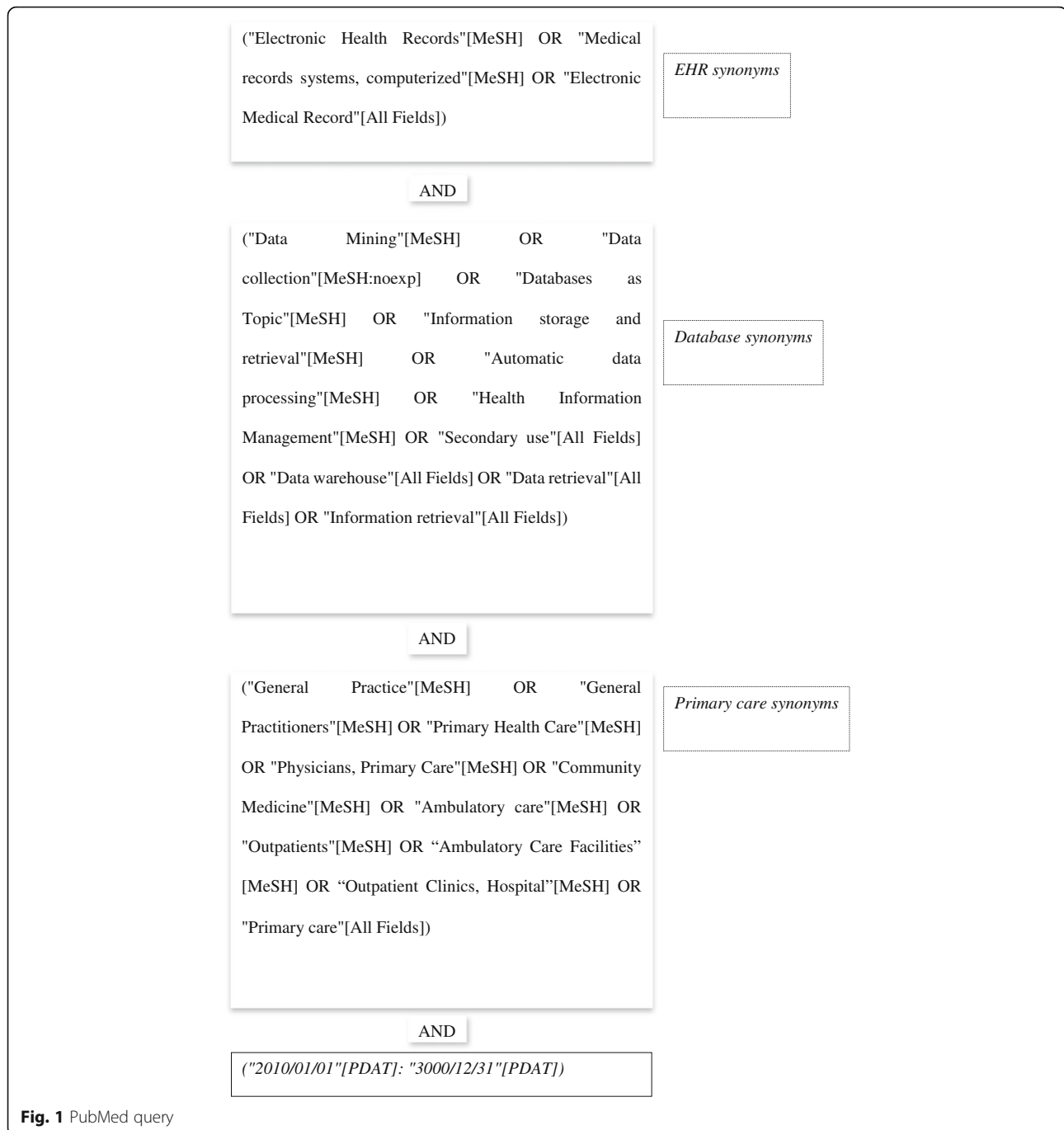
Secondary to our PubMed search, we identified 36 routine data collection projects. Our screening of the gray literature consists of the screening of each project website. First, we tried to find the official websites from our literature review. Then, when the website was not quoted in the references of the paper, we used the Google search engine. The name of the database was associated with the keywords "database" or "primary care database" enclosed in quotation marks.

Moreover, to identify the project managers, the project contact person quoted in the website was contacted by email, when such information was present in the article/website.

Study selection

Inclusion criteria

Articles were retained if they referred exclusively to EHRs (as opposed to paper-based health records), focused on automatic data processing (rather than on



manual data analysis) and the EHRs were retrieved from primary care databases. Secondary or tertiary care data were not considered. Within primary care data, raw patient medical records were included, while registries, which contain already processed data, were excluded. Databases containing primary and secondary care data were included only if the study concerned the primary care population. An article was considered to be a routine primary care data

collection project when it analyzed EHR data from different GPs.

Exclusion criteria

Articles published before 2010 were excluded due to the introduction of new MeSH terms, for instance "Electronic Health Record", in 2010. The 2010 cutoff was indeed primarily used due to the introduction of more relevant MeSH headings in 2010 so as to define active

databases. It allowed us to retrieve active routine data collection projects, because they published papers recently and our PubMed query was more accurate thanks to the MeSH term indexing.

Moreover, we excluded:

- Articles not written in French nor English languages
- Articles on projects not meeting the inclusion criteria
- Articles that could not be retrieved as full text, due to an absence of subscription to the review
- Articles in which the original database was not precisely identified.

Data collection

The PubMed search query was launched independently by two of the authors who then read the abstracts to select relevant publications on the basis of the inclusion and exclusion criteria. They then independently read the full text of the retained articles to confirm that the inclusion and exclusion criteria were met. Disagreements were solved by consensus.

This allowed us to identify the major primary care data collection projects and to compare their prevalence worldwide. Each PubMed article was screened to retrieve information on the data collection project and its stakeholders with the objective to compare projects. Data were retrieved from the included studies using a standardized form (Fig. 2). This form was based on that of a Canadian qualitative study on the use of primary healthcare EHRs for research [14], and was revised by the authors during a working discussion. Indeed, this paper allowed us to extrapolate three parts of our form: technological infrastructure, GPs' roles, data collection network stakeholders. The form was pre-tested on 5% of the selected papers and was not modified after this test. Two authors read separately seven articles and filled in the form. Then, they compared their results. There were no disagreements due to the objective nature of the data extracted. Consequently, the form was left unmodified.

Each official project website also was screened using the same form. Moreover, each contact person received one single e-mail message to gather information about the data collection project partners (identification of the involved parties).

Three groups of stakeholders were defined from the various partners retrieved in papers and websites: governmental services, academic institutions, software companies, so as to represent each identified extracted partnership. The definition of stakeholders groups was ascertained by [12], for instance we can quote "Within this study, we define stakeholders as those individuals holding an interest in the topic of EMRs in PHC; these

1 -Project features

- Name
- Country
- Number of patients included
- Number of GPs
- Geographical location of the network

2 – Technological infrastructures

- Name of the used software
- Extraction tool
- Comprehensive IT platform
- Data linkage tools

3 – GPs' role

- Place of GPs in the primary care organization
- GP incentives

4 – Data collection network organization

- Governance, partnerships
- Funding

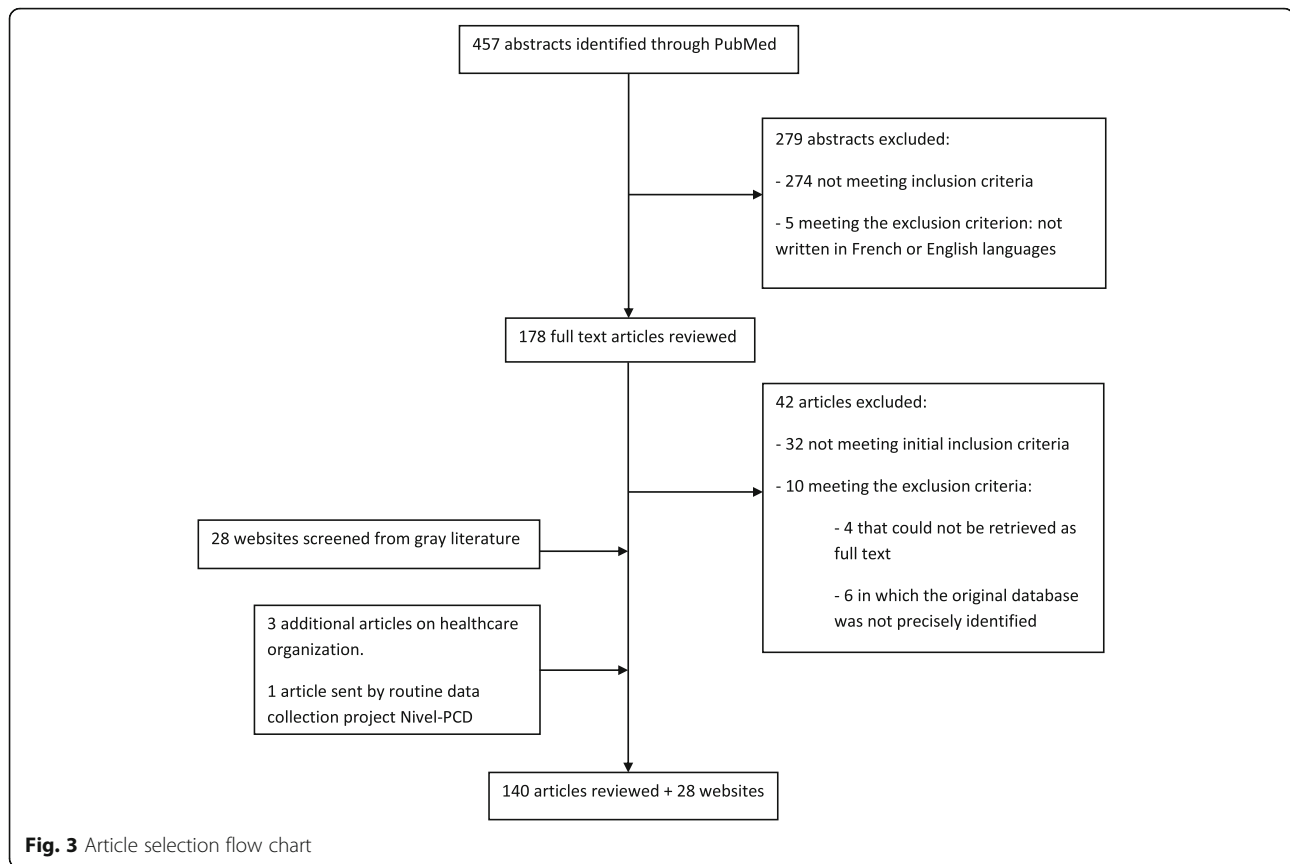
Fig. 2 Form used to analyze the selected primary care data collection projects and relevant websites

individuals included clinicians/healthcare practitioners, decision-makers (those who make policy and health planning decisions), researchers and EMR vendors."

Results

Selection of articles on primary care data collection projects

The PubMed search yielded 457 article abstracts among which 279 were not retained based on the inclusion/exclusion criteria (Fig. 3). The remaining 178 articles dealt with primary care routine data collection projects. Another 42 articles were excluded based on the inclusion and exclusion criteria. Three articles relevant for understanding the healthcare organization and one paper sent by Nivel-Primary Care Database (Nivel-PCD) were added for the final analysis. Each contact person from the 28 websites identified by screening the gray literature was emailed to ask information about the data collection project stakeholders. The Information System for the Development of Primary Care Research (SIDIAP) and Canning Division of General Practice contacts were not operational. We received ten answers and no follow-up email message was sent.



Comparison of the selected routine primary care data collection projects

List of routine primary care data collection projects

Review of the retained articles allowed the identification of 36 projects on collection of data from primary care EHRs (Table 1). They were mainly from English-speaking countries (USA, UK, Canada, Australia) as shown by Fig. 4, but also from various European countries. The TRANS-FoRm project involved different European countries.

By analyzing the percentage of retrieved publications (i.e., the number of papers per country extracted in our query divided by the total number of papers), countries could be classified in four groups (Fig. 5). The first group included the UK that had the highest percentage of publications compared with all the other countries (40% of all retrieved publications). Canada (19%) and the United States of America (USA) (18%) formed the second group. The third group included few European countries (Netherlands, Spain, Belgium and Italy) and Australia. These countries had successful primary data collection projects, although they were not published on as much as the other groups (2 to 8.5% of all retrieved publications). The last group comprised four European countries (France, Malta, Sweden and Switzerland) that had very few primary data collection networks.

Analysis of the publications linked to a specific primary data collection project within a country showed that the UK hosted several projects, among which CPRD (38 papers retrieved) and The Health Improvement Network (THIN) (12 papers retrieved) were the most used (Table 1). Moreover, according to their official websites, research using CPRD and THIN data has led to more than 1500 and 500 publications [15, 16], respectively. In Canada, the most quoted were the Canadian Primary Care Sentinel Surveillance Network (CPCSSN), which includes the Deliver Primary Healthcare Information (DELPHI) program, and the Electronic Medical Record Administrative Data Linked Database (EMRALD) project. In the USA, the Veterans Health Administration was the main provider of publications. In continental Europe, the Integrated Primary Care Information Project (IPCI) and Nivel-PCD in the Netherlands were associated with a significantly higher number of publications compared with other European projects. The Spanish SIDIAP also was quoted several times. In Italy, all projects were managed by the same private-sector company: Cegecim Strategic Data©. Australia had several local and apparently independent initiatives. Belgium had two databases: Intego and the National Scale Routine Data Collection Network (NHIDI).

Table 1 List of the routine primary care data collection projects

Project name or network associated with the database	Country of origin	PubMed occurrence (ref)	Official website (ref)
Geographical and Resource Analysis in Primary Health Care (GRAPHIC)	Australia	1 [48]	[49]
Canning Division of General Practice in Western Australia project	Australia	1 [50]	No website
Melbourne East General Practice Network in Victoria project (MEGPN)	Australia	1 [17]	[27]
Local initiative: three general practices	Australia	1 [51]	No website
Intego Project	Belgium	2 [52, 53]	[54]
National Scale Routine Data Collection Network (NIHDI)	Belgium	1 [23]	No website
Canadian Primary Care Sentinel Surveillance Network (CPCSSN)	Canada	13 [36, 55–66]	[24]
Deliver Primary Healthcare Information (DELPHI)	Canada	2 [25, 67]	[24]
Electronic Medical Record Administrative Data Linked Database (EMRALD)	Canada	10 [34, 35, 37, 62, 68–73]	[74]
One community health center – local experiment	Canada	1 [18]	No website
Longitudinal Patient Data Network - Cegedim®	France	1 [8]	Commercial website
Cegedim® Strategic Data - Longitudinal Patient Database (CSD)	Italy	3 [38, 75, 76]	[77]
TransHIS - TRANSFoRm Project - Translational Research and Patient Safety in Europe	Malta	1 [78]	[79]
TransHIS – TRANSFoRm Project - Translational Research and Patient Safety in Europe	Netherlands	1 [78]	[79]
Integrated Primary Care Information Project (IPCI)	Netherlands	5 [38, 76, 80–82]	[29]
Nivel: Primary Care Database (NIVEL-PCD) and LINH Database (Netherlands Information Network of General Practice)	Netherlands	4 [83–86]	[41]
Julius Clinical Primary Care Research Network (JPCRN)	Netherlands	1 [87]	[39]
Registration Network Family Practices (RNH)	Netherlands	1 [88]	[89]
SIDIAP, Information System for the Development of Primary Care Research	Spain	3 [90–92]	[92]
Electronic Clinical Record in Primary Care from Madrid	Spain	1 [93]	No website
BIFAP, Database for Pharmacoepidemiology	Spain	1 [94]	[95] – (website excluded because written in Spanish)
Swedestar, EHR system for primary care	Sweden	1 [96]	No website
Family Medicine ICPC-Research using Electronic Medical Records (FIRE project)	Switzerland	1 [97]	[98] –(website excluded because written in German)
Clinical Practice Research Datalink (CPRD) or General Practice Research Database (GPRD)	United Kingdom	38 [10, 40, 99–134]	[15]
The Health Improvement Network (THIN)	United Kingdom	12 [11, 135–145]	[16]
Q Research	United Kingdom	1 [11]	[42]
Local initiatives or uncompleted project, i.e. General Practice Extraction Service	United Kingdom	6 [146–151]	No website
Veterans Administration – Corporate Data Warehouse	United States- <i>Several states</i>	8 [152–159]	[28]
Geisinger® Health System Project	United States- <i>Several states</i>	3 [160–162]	[30]
General Electric® Centricity Electronic Medical Records Research Database	United States- <i>Several states</i>	2 [163, 164]	Commercial website
OCHIN, Inc. Oregon Community Health Information Network	United States - <i>Several states</i>	1 [22]	[165]
Baylor Health Care System (BHCS) and Christiana Care Health System (CCHS)	United States	1 [166]	[167]
University of California Davis Health System Electronic Medical Record System	United States – <i>California</i>	1 [168]	[169]
CHCI, Community Health Center, Inc.	United States – <i>Connecticut</i>	1 [170]	[171]
Massachusetts General Physicians Organization, MGPO	United States – <i>Massachusetts</i>	1 [19]	[172]
4-site federally qualified CHC, Open Door Family Medical Centers (Open Door), located in New York	United States – <i>New York</i>	1 [173]	[174]

Table 1 List of the routine primary care data collection projects (*Continued*)

Project name or network associated with the database	Country of origin	PubMed occurrence (ref)	Official website (ref)
New York City Primary Care Information Project	United States – <i>New York</i>	1 [21]	[175]
Institute for Family Health (IFH)	United States – <i>New York area</i>	1 [20]	[176]
Isolated initiative: Dallas (Texas)	United States – <i>Texas</i>	1 [177]	No website
University of Wisconsin Department of Family Medicine Clinical Data Warehouse	United States – <i>Wisconsin</i>	2 [178, 179]	Unavailable website
University of Wisconsin Electronic Health Record - Public Health Information Exchange (UW eHealth - PHINEX)		1 [33]	

This analysis allowed identifying major routine primary care data collection projects: CPRD, THIN, the Veterans Affairs Corporate Data Warehouse, CPCSSN and EMERALD.

Patients and healthcare professionals' coverage

The number of patients in the database and the number of involved GPs or practices (two parameters that are representative of the data collection project scale) are summarized in Table 2. Primary care data collection projects in the UK, Netherlands, Spain and Switzerland included the highest numbers of patients as represented in Fig. 6.

The number of participating GPs varied from 30 to 7200, depending on the country or project. This information was not retrievable for 17 projects. The investigators were often described as a group of healthcare professionals.

Most primary care data collection projects were implemented nationwide and not limited to a specific geographical location within a country. Nevertheless, several networks were based on location similarity, for instance the Melbourne East General Practice Network in Victoria [17], community health centers in Canada [18] or in the USA (for instance, greater Boston area [19] and New York [20, 21]). The EMERALD and DELPHI projects originated from Ontario, a Canadian province.

Technological infrastructure of the IT systems used for data collection and data reuse

Technological infrastructures are key elements of routine data collection and reuse projects.

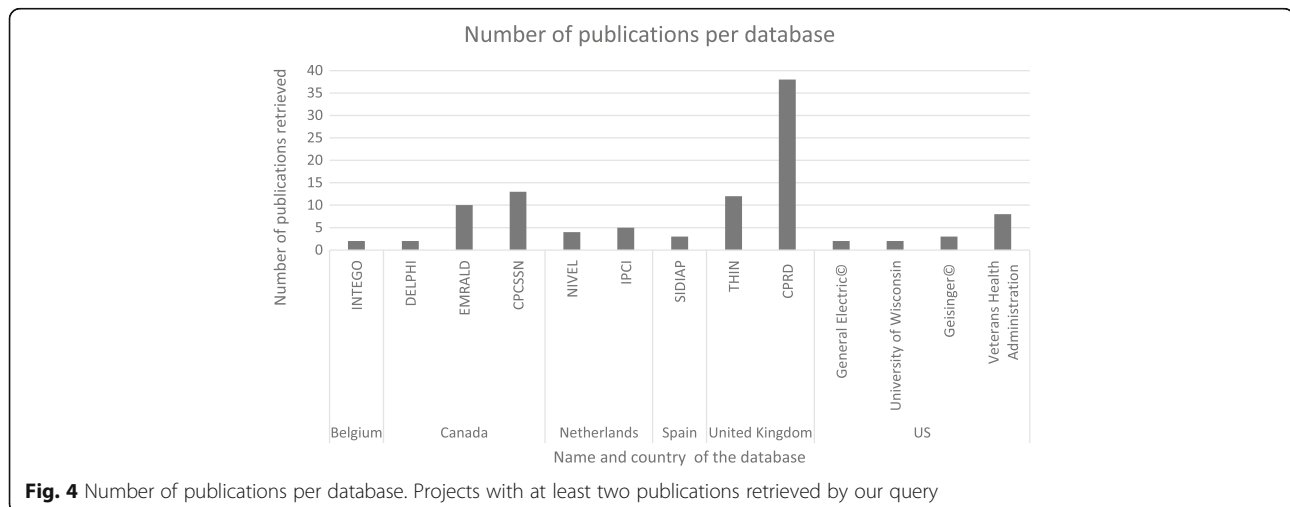
We distinguished these four items:

- EHR software with extraction tools
- Data warehouse, which import, classify and store data
- Functional integrated platforms, which gather tools to exploit data coming from the data warehouses
- Data linkage facilities, which give the possibility to cross data with other databases

They represent different levels of achievement of data collection and reuse projects.

EHR software tool as source of data collection

Most projects were set up in collaboration with a single EHR software company and a single IT system or software tool was used to initiate the routine primary care data collection project (Table 3). In the USA, the Veterans Health Administration was especially praised for developing an open source electronic medical records system: the Veterans Information Systems and Technology Architecture (VISTA). The choice of EHR vendor and the negotiations for the initial EHR software



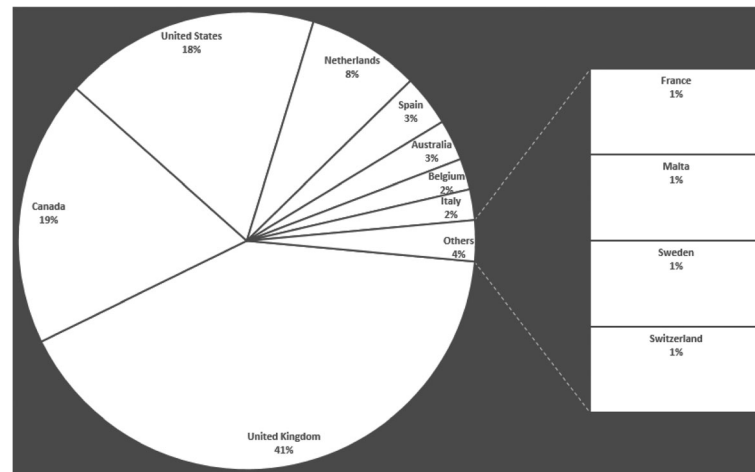


Fig. 5 Chart representing the percentage of publications per country

purchase were considered as the first major challenges in building the OCHIN database [22]. Moreover, the availability of many different software applications seems to hinder data collection. For instance, in Belgium, more than 17 different software systems are currently used by GPs and this hampered the development of primary care data collection projects, according to the authors [23]. Indeed, the use of different software applications increases the data collection complexity and adds interoperability issues. Nevertheless, some projects managed to collect data from different software systems with various database architectures. For instance, in the CPRD, primary care observational data came from three main EHR systems for GPs. However, historically, all data were retrieved from the Vision software [15]. For the CPCSSN, data were extracted from ten EHR systems [24]. The DELPHI program originally collected data only from the EHR software Healthscreen [25], but now contains data from various EHR software systems [24].

Software development companies, which produce the EHR software, were usually requested to implement the data extraction tools. In Belgium, EHR software developers were asked to develop an extraction module to support NIHDI uploading procedure [23]. In Switzerland [26], the companies that provided EHR software for the FIRE project were asked to update their products with an exporting tool (called Ghranite) to enable the automatic downloading of core FIRE variables from individual EHRs to a central server [27]. The OCHIN project also stressed the importance of finding a software vendor who was willing to provide system modifications and enhancements.

Finally, offering simplified data extraction tools to minimize the additional workload is a major factor. Otherwise, the applicability and utility of EHR data for large-scale research purposes remains limited as pointed out by most of the project websites.

Implementation of comprehensive data reuse platforms

The technical architecture of the IT system for data reuse was poorly detailed in all articles. However, we could extract some trends concerning the delivery of data mining solutions.

Considering the aim of routine data collection projects is to analyze data, this part is essential to expand data exploitation and bring data reuse to a higher level of use.

A centralized official data warehouse A data warehouse has several roles: it imports, classifies and stores data coming from the EHR software tool.

In most cases, the data warehouse was linked to an official administration (university, national administration warehouse) that brought the added value of official recognition. For instance, in the Q research project, the EMIS software transmitted in a secure way all aggregated data to the University of Nottingham that was and still is the only access point to the full database. In the USA, the Veterans Administration built the Corporate Data Warehouse and four Regional Data Warehouses to provide a standard architecture that centralizes all clinical data [28]. In the Netherlands, most GP databases, including IPCI, are connected to one of the eight national Medical University Centers [29].

A functional integrated platform: A turnkey service

We define as functional integrated platforms, services which gather tools to exploit data coming from the data warehouse.

According to the official websites of the major primary care data collection projects identified in this analysis (i.e., CPRD, THIN, Veterans Administration database and CPCSSN), they facilitate data exploitation by providing networks that extract datasets from the data warehouse for researchers and by offering a range of services

Table 2 Demographics of the primary care data collection projects

Country	Project name	Number of patients (% population) ^a	Number of GPs (other healthcare measurement)	Sources
UK	CPRD	13 million (20%)	600 GPs	[15]
	THIN	11.1 million (17%)	562 GPs	[16]
	Qresearch	18 million (28%)	1000 GPs	[42]
USA	Veterans database	17 million (5%)	GPs unknown (53,000 healthcare professionals)	[154]
	Geisinger© Health System project	400,000 (0.1%)	GPs unknown (41 community practice clinics)	
	General Electric Centricity©	8.9 million (2.8%)	7259 GPs	[163, 164]
	OCHIN, Inc	1 million patients annually (0.3%)	GPs unknown (200 clinic sites)	[22]
	Baylor Health Care System (BHCS) and Christiana Care Health System (CCHS)	680,000 patients (0.2%)	GPs unknown (33 primary care practices + 35 office practices)	[166]
	University of California Davis Health System's electronic medical record system	unknown	GPs unknown (13 clinics in this analysis)	[168]
	CHCI – Connecticut	130,000 (0.04%)	GPs unknown (13 primary health centers)	[170]
	MGPO	87,568 patients in this study (0.02%)	148 GPs in this study	[19]
	OpenDoor	40,000 patients annually (0.01%)	GPs unknown (4-site federally qualified community health center in New York)	[173]
	New York City Primary Care Information Project	unknown	GPs unknown (82 practices)	[21]
	IFH	90,000 patients (0.02%)	GPs unknown (17 community health centers)	[20]
	Department of Family Medicine – Wisconsin	110,000 for these studies (0.03%)	GPs unknown (28 primary care clinics)	[178, 179]
Canada	CPCSSN	1.28 million (3.6%)	1031 GPs	[24, 56]
	EMRALD	0.5 million (1.4%)	350 GPs	[74]
	DELPHI	30,000 (0.08%)	30 GPs	[24]
Netherlands	IPCI	1.5 million (8.8%)	600 GPs	[29, 38]
	Nivel-PCD	1.6 million (10%)	GPs unknown (500 practices)	[41, 83, 85]
	LINH	350,000 (2%)	85–150 GPs	
	Julius	156,176 (0.9%)	60 GPs	[41, 87]
	RNH	51,700 (0.3%)	65 GPs (22 practices)	[89]
Spain	SIDIAP	6 million (12.8%)	274 GPs	[90]
	ECRPC	unknown	GPs unknown	
	BIFAP	3.2–4.8 million (6.8–10%)	1183 GPs	[94, 95]
Australia	GRAPHIC	32,000 (0.1%)	GPs unknown (1 GP practice)	[48]
	Canning Division of General Practice	unknown	GPs unknown (8 GP sites)	[50]
	MEGPN	505,600 (2.1%)	578 GPs (141 practices)	[17]
Belgium	Intego	285,000 (2.5%)	GPs unknown (95 GP practices)	[52]
	NIHDI	29,000 (0.25%)	GPs unknown (5000 practices)	[23]
France	LPD – Cegedim	1.5 million (2%)	1200 GPs (+ 750 various medical specialists)	[8]
Italy	Health Search Database / CSD©	2.5 million (4.1%)	900 GPs	[77]
Sweden	Swedestar project	unknown	GPs unknown	[96]
Switzerland	FIRE	113,335–293,000 (14–36%)	60–120 GPs	[97, 98]

^aPercentage denominator (i.e., the country population was extracted from national official websites)

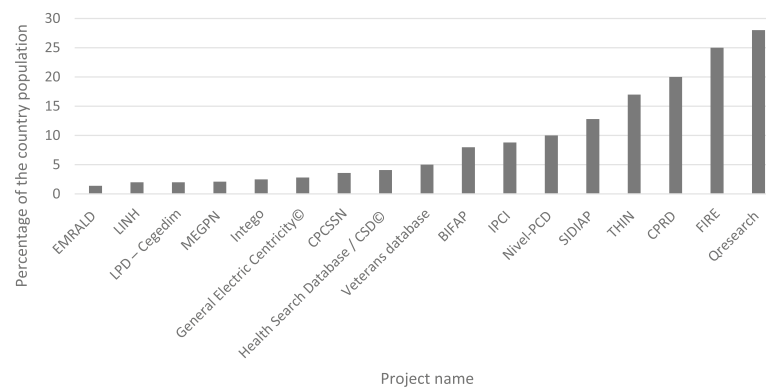


Fig. 6 Representativeness of population in projects. Projects collecting data on more than one percent of the country population

and products in the areas of medical research and public health care. For instance, in the USA, data retrieved from healthcare facilities have been made available for research purposes in a data warehouse (corporate data warehouse). On top of this warehouse, standard tools were developed for robust access, reporting and data analysis at the enterprise level. The Veterans Affairs Informatics and Computing Infrastructure (VINCI) is an initiative to improve the researchers' access to Veterans Affairs data and to facilitate their analysis, while ensuring data privacy and security [28].

Besides these global extraction systems, few modules have been developed with specific functionalities related to research. For instance, the CPRD software system is integrated with EHR systems to enable randomization at point of care in a real world setting [15].

Integrating data linkage functionalities Data linkage facilities give the possibility to cross data with other databases. To expand the data processing potential, linkage with other databases has been included in the integrated IT systems used for these projects. CPRD services were progressively developed to increase the primary care data coverage and the number of linked datasets [15, 16]. For instance, CPRD GOLD, the primary care database of CPRD, has access to the following linked datasets: Death Registration data from the Office for National Statistics, Cancer Registration data from Public Health England, Hospital Episodes Statistics data. The Geisinger® Health System warehouse receives feeds from multiple-source systems, including data from EHRs and also data on financial decision support, claims and patient satisfaction [30]. SIDIAP allows linking to other Catalonia databases [31], such as CMBD-AH (dates, diagnoses and procedures linked to admissions in each Catalonia hospitals) and the death registration database (date and causes of death of all Catalonia residents). Globally, data linkage with other databases has been developed by many IT systems to expand the potential of their network. Moreover, as highlighted by

Phillips and colleagues, “Integrated big data systems could also collect information directly from patients about their health behaviors, community resource utilization, social networks, and other social determinants of health” [32]. Similarly, the University of Wisconsin Electronic Health Record-Public Health Information Exchange (UW eHealth-PHINEX) can be used to study “health and disease within the patient’s biologic, psychosocioeconomic, environmental, and community context” [33]. In this database, EHR data can be linked to geographic, environmental, socioeconomic, and demographic statistical data to obtain comprehensive information for the investigation of infectious, acute, chronic, injury-related, occupational and environmental health outcomes and risk factors. Then, it allows using multivariate analysis and data mining tools to identify variables for predicting disease and health quality at the census block level.

General practitioners' involvement in the development of data collection networks

GPs as care pathway coordinators

Gate-keeping is the identification of patients with one primary care provider who is most responsible for their care. Gate-keeping by GPs has been recently adopted in most European countries and in many countries worldwide. The inherent central place accorded to GPs enhances their key role in “leading primary care and community big data efforts” (172). The GPs' key role in the development of data collection networks was particularly highlighted in the Canadian primary care data collection projects. In Ontario, following primary care reforms, patient rostering (i.e., the connection of one patient to one physician or physician team) has been introduced [34] and more than 80% of the population is rostered to a primary care physician [35]. Conversely, other Canadian provinces do not currently have similar broadly based primary care patients' enrollment systems [36]. Analysis of the retained articles indicated that

Table 3 Stakeholders

	Governmental services	Academic institutions	Software companies Private-sector companies
CPRD ^a	- Medicines and Healthcare products Regulatory Agency (MHRA) - Department of health via the Health and Social Care Information Centre (HSCIC) [15]	National Institute for Health Research (NIHR)	IMS Health [®] [180] Several EHR software systems [15]
THIN			In Practice System - Cegecim [®] , IMS Health [®] [16]
Q research		University of Nottingham and its division of Primary Care [42]	EMIS [42]
Veterans database ^a	Veterans Health Administration [152]		VISTA, developed by the Veterans Health Administration [152]
Geisinger [®] Health System project			Geisinger [®] Health system [30] IBM [162]
General Electric Centricity [®] electronic medical records research database			General Electric [®] [181]
OCHIN, Inc	US Agency for Healthcare Research and Quality Funding from the American Recovery and Reinvestment Act	OCHIN SafetyNetWest Practice-Based Research Network, non-profit partnership with academic researchers, governed by a board of directors	EpicCare [22]
Baylor Health Care System (BHCS) and Christiana Care Health System (CCHS)	Agency for Healthcare Research and Quality, US Department of Health and Human Services		One commercial EHR system [166]
University of California Davis Health System electronic medical record system		University of California [168]	University of California Davis Health System electronic medical record (Epic, Verona, WI)
CHCI			eClinicalWorks [170]
MGPO		Massachusetts General Hospital - Radiology Department associated with the Massachusetts General Physician Organization Primary Care Operations Improvement Group	
OpenDoor		Columbia University New York University School of Medicine Division of General Internal Medicine Mailman School of Public Health, Columbia University Primary Care Development Corporation	eClinicalWorks [173]
New York City Primary Care Information Project	Agency for Healthcare Research and Quality		eClinicalWorks [21]
IFH	Funding from New York state Health Resources and Services Administration		EpicCare ambulatory EHR [20, 176]
Department of Family medicine - Wisconsin	Wisconsin State Division of Public Health	Department of Family Medicine University of Wisconsin [178]	
CPCSSN ^a and Delphi	Public Health Agency of Canada [24]	- Centre for Studies in Primary Care at Queen's University [24] and several universities - College of family physicians [24]	Several EHR software systems [24]

Table 3 Stakeholders (Continued)

	Governmental services	Academic institutions	Software companies Private-sector companies
EMRALD ^a	Ontario Ministry of Health and Long-Term Care, depending on the projects [74]	<ul style="list-style-type: none"> - Institute for Clinical Evaluative Sciences, a prescribed entity under Ontario's Personal Health Information Protection Act [70] - University of Toronto and its Department of Family and Community Medicine - University of Ottawa 	Several EHR software systems: Practice Solutions Software electronic medical record, Oscar [74]
IPCI ^a	No	Erasmus University Medical Center Rotterdam: Department of Medical Informatics, Unit of Pharmaco-epidemiology [29], Department of General Practice [29]	No
Nivel-PCD ^a	Dutch Ministry of Health [41]	<ul style="list-style-type: none"> - Groningen University, Free University of Amsterdam, Utrecht University and more - National college of GPs - National association of GPs [41] 	NIVEL owns the Information Communications Technology (ICT) department and has close links with vendors of ICT systems [41]
LINH	Netherlands Institute for Health Services Research (NIVEL) Dutch Ministry of Health [41]	<ul style="list-style-type: none"> - IQ healthcare - National association of GPs - Dutch College of GPs [41] 	
Julius General Practitioners Network (JGPN) database ^a	No	<ul style="list-style-type: none"> - Julius Center for Health Sciences and Primary Care at Utrecht University Medical Center [39]. - GPs' Networks at each Dutch University - Representatives of care groups (groups of GPs from the same area) have a seat in the scientific committee. 	Information Communications Technology company (called Proigia) extracts primary care data from GP practices and send them to the database
RNH		Department of Family Medicine, subdivision of the Faculty of Health, Medicine and Life Sciences at Maastricht University. Maastricht University Medical Center. CAPHRI, school for public health and primary care [89]	
Transform project ^a	Depending on the country	Several universities worldwide (e.g., Crete, Antwerp) [79]	TransHIS Custodix, Belgium, for central data linkage
SIDIAP	Catalan Institute of Health [31]	<ul style="list-style-type: none"> - Jordi Gol Primary Care Research Institute - Girona Research Support Unit [31] 	eCap, Catalan software company [31] IMS Health® [180]
ECRPC Madrid	Madrid Regional Health Authority		
BIFAP ^a	6 projects administered and financed by Agencia Española de Medicamentos y Productos Sanitarios (Spanish Agency of Medicines and Medical Devices), a public agency belonging to the Spanish Department of Health [95] - Institutional collaboration of the regional governments of the nine participant Autonomous Communities	<ul style="list-style-type: none"> - Department of General Practice - Learned societies of GPs' support - One GP who is an active BIFAP collaborator appointed as member of the BIFAP Scientific Committee. 	in-house IT and pharmaco-epidemiologist teams No private-sector company
GRAPHIC		<ul style="list-style-type: none"> - Australian National University - Australian Primary Health Care Research Institute [182] 	

Table 3 Stakeholders (Continued)

	Governmental services	Academic institutions	Software companies Private-sector companies
Canning Division of General Practice		Australian General Practice Network	
MEGPN	Independent accredited not-for-profit organization [27]		PCS Clinical Audit Tool® Practice Health Atlas®
Intego ^a	Flemish government, Department of Health Care [54]	- Department of General Practice of the KU Leuven [54] - Other universities. - GPs' participation in the ethics and scientific committee	- Datasoft, Medidoc - Careconnect EHR - No private-sector company involved in data analysis
NIHDI	Public social security institution [23]		Extraction module on EHR [23]
Health Search Database / CSD®			Cegedim® [77]
Swedestar project		- Department of Medical and Health Sciences, University of Linköping, Unit of Research and Development in Primary Care, Jönköping - Lindsdals Primary Health Centre [96]	Swedestar software [96]
FIRE		- Institute of General Practice of Zurich University [97] - "SGAM informatics", a study group of the Swiss Association of GPs [26]	Grhanite extracting tool
LPD network - Cegedim			Doc'ware software - Cegedim® [8]

^athe contact person gave information on the project governance following our enquiry by email

primary care data collection networks (DELPHI, EMRALD) are most developed in Ontario.

Role of the type of primary care organization

Analysis of the selected articles indicated that primary healthcare organizations differ among countries, from GP's office to outpatient clinics. However, the type of primary care organization (for instance, small GP practices in the UK and outpatient clinics in the USA) did not seem to influence the development of successful primary care data collection projects.

GPs' contribution and incentives

GPs voluntarily agree to supply data to the data collection project. Sometimes, their contribution required high quality and completeness standards [16, 37], coding accuracy [38], or seniority in software use [34, 37]. Moreover, incentives may be offered to improve GPs' contribution and compliance. For instance, data providers receive a percentage of the profits generated by research carried out using THIN [16]. The Julius network also compensates financially the time spent in supplying data [39]. In addition, GP practices involved in the CPRD project can generate research revenues through their involvement in studies that require validation, sample collection, or patient questionnaires [15]. Similarly, the THIN project offers opportunities for extra

payments to GPs when researchers require supplementary data about selected patients [16]. Conversely, Q research is a not-for-profit network and does not have funding to compensate GPs for their contribution [40].

Software training sessions also have been proposed to GPs involved in data collection for the THIN or Julius network [16, 39, 41]. Moreover, regular feedback and reports on data recording might be given to the participating GPs [16, 39, 41]. Indeed, GPs are increasingly interested in using their EHR data to better understand and manage their patient populations. GPs are also encouraged to start their own research project [42].

Stakeholders of routine primary care data collection projects

The contact information of 36 primary care data collection projects was used to enquire about the project managers by email. Based on their responses ($n = 10$) and the information found in the articles or websites, we were allowed identifying three major project governance actors:

- governmental services (Veterans Health Administration in the USA, National Health Service in the UK)
- academic institutions (universities and department of general practice) and GP representatives

(Nottingham University, Department of General Practice of Leuven University, Dutch College of General Practitioners)

- private-sector companies specialized in software development (Cegedim[®]) or in data extraction and analysis (IMS Health[®])

The collected information highlighted that only few projects were exclusively managed by software companies. IMS Health[®], which bought Cegedim[®] Customer Relationship Management Software and Strategic Data in 2015, is the leader in several countries. We noted that the partnership between CPRD and IMS Health[®] was not mentioned in the CPRD website or in the reply following our enquiry by email, but was quoted in the IMS Health[®] website.

On the other hand, most projects were supported by academic or governmental institutions in association with software companies (particularly for the extraction process). Generally, universities were involved via their general practice or medical informatics departments. Governmental services also were involved, but their role (funding, ethical validation) was often not specified. Concerning governmental services, a unique official initiative is the Health Information Technology for Economic and Clinical Health (HITECH) Act in the USA that defines the direct involvement of the state in the development of data reuse. More precisely, it supports the meaningful use of certified EHR technology, “connected in a manner that provides for the electronic exchange of health information to improve the quality of care; and provides to the Secretary of Health & Human Services (HHS) information on quality of care and other measures” [43].

Discussion

This review of the literature allowed the identification of determinants that favor the development of durable routine data collection projects.

Methodology

Study strengths

This is the first review that lists the databases derived from primary care electronic health records and their organizational features.

Moreover, we followed the PRISMA criteria [12], an evidence-based minimum set of items for reporting in systematic reviews. Then, quality was controlled using the AMSTAR checklist [13].

Study limitations

Methodological issues The aim of this study was to identify the determinants within the healthcare organization that influence routine data collection

projects in primary care. We could identify 36 projects by screening the PubMed database. We only used this database, but added information from the gray literature.

Missing or partial data We took for granted that the PubMed occurrences were correlated with the international outreach of primary care databases. This hypothesis was validated by the fact that our query identified major routine primary care data collection projects.

Nevertheless, there is actually a discrepancy between the effective number of publications of each project and the number of publications retrieved in our paper. There are several reasons to explain the differences.

- Firstly, our query search filters by the publication year. CPRD database was created in the 1980s, so numerous papers are not taken into account by our query.
- Secondly, publications listed on the websites are not indexed into PubMed.
- Thirdly, due to the lack of MeSH term to index primary care databases, our query search was not able to retrieve all papers published. Indeed, if we analyze the PubMed keywords of articles that were not returned by our query but which were recorded on database websites, you can note that they are not referenced as primary care data collection projects, but as the main topic of the article. For instance, the paper “Prevalence, incidence, indication, and choice of antidepressants in patients with and without chronic kidney disease: a matched cohort study in UK Clinical Practice Research Datalink” is indexed with the keywords “antidepressants; chronic kidney disease; depression; incidence; prevalence”.

Finally, our initial purpose was to find papers describing the characteristics of the data mining projects and not all the results of the exploitations of such database. Including all the publications was not achievable because all websites do not reference their publications and it was not answering our initial methodology. Concerning the project description, the number of patients was difficult to extract from the articles because several studies were based on subsets of the available databases. We finally chose to take into account, when possible, the figures indicated in the website because it described the whole project. It would have been interesting to compare projects according the patient-year data, but this figure was rarely available.

Another issue was the number of GPs per project. Indeed, the method of representation of healthcare professionals varied in the different countries (e.g., number of GPs, number of healthcare professionals, number of community practices). We observed that the number of

patients was not proportional to that of GPs, suggesting that the GPs' involvement varied in function of the project nature or the primary care organization of the specific country. Moreover, the number of participating GPs could be different according to the chosen data mart [23].

There was no systematic correlation found between the representativeness of population in the database, the number of GPs and the number of publications per database. We thought that the larger the database is, the more publications are edited. But, the amount of data can also be an hindrance and without comprehensive data formatting, data mining turns out to be more complex. This result could invite us to begin with the creation of a local routine data collection project before expanding it to a larger scale in our country.

We reviewed each article and website using a predefined form (Fig. 2). Nevertheless, some data were missing especially about stakeholders.

Finally, this study focused on successful data collection projects and therefore we could not identify factors that could limit GPs' participation, such as privacy issues, lack of training and information.

Summary of evidence

Geographical distribution of routine primary data collection projects

This study highlights the leadership of English-speaking countries: UK, USA and Ontario, in Canada, are clearly ahead of European countries.

Moreover, geographical proximity was sometimes important for setting up and implementing a project, for instance the EMERALD project. Similarly, in France, hospitals from the same area often pool together their information to aggregate a larger database [3, 4], and routine primary data collection projects and initiatives are primarily local (only one national project) [3, 9]. Nevertheless, geographical proximity is then pushed into the background by the virtual nature of electronic information. Thus, the network effect seems to be a facilitating, but not a determining factor.

The major role of EHR software in data extraction/collection

When it comes to the implementation of a routine primary data collection network, software companies play a key role directly via their software system and the development of data extracting tools. Most of the reviewed projects included at least one software company, suggesting that they are part of the foundations of such projects.

Moreover, most projects were linked to a single software application, thus limiting interoperability issues and technically facilitating data analysis. This suggests that the number of software systems hinders the development of a primary care data research networks.

Indeed, interoperability issues hinders the development of a unique database. Data sources come from EHR and the presentation of data can be heterogeneous considering each software application has its own logic and data schemas. Data are not identically formatted and thus are not always comparable. In France, more than 12 software applications are used [44] and this could explain the difficulty to create a nation-wide primary care data collection network.

A comprehensive IT system dedicated to data exploitation

Our study also highlights that data must be processed and integrated before being released to final users, generally academic researchers. Projects often offer support to researchers for data extraction and analysis. The provision of data sources (EHR data) requires formatting and processing before being included in the data warehouse. Then, a second step consists of the extraction of datasets from the data warehouse. Enabling users to shuttle data sets of any type seems difficult considering the large amount and complexity of data. Indeed, the big data area requires a deep knowledge of the content and form of the data handled.

This implies the development of a comprehensive IT platform to provide services and tools to end users. In France, we combined secondary care data with administrative information from the National Health Insurance Cross-Scheme Information System. This led to the development of a global IT system that includes a data management platform where data are available to researchers via dedicated tools [2].

GPs: Care pathway coordinators who are essential for data collection

Factors enabling GPs involvement The organization of primary care varies depending on the countries (from isolated GPs to outpatient clinics), but this does not seem to affect the development of primary care data collection projects. Particularly, routine primary data collection projects are not more numerous in countries with big primary care facilities. Nevertheless, a local network effect can occur and facilitate the spreading of such initiatives.

In addition, several strategies are often put in place to increase and optimize GPs contribution. GPs participate voluntarily, but their participation can be promoted by the various advantages they can have in return: financial benefits, training sessions, regular feedback, and participation in research programs. These benefits could be proposed also to French GPs to create a French primary care data collection project.

Training sessions for GPs included data coding training because most of the data are coded using International

Classification of Primary Care version 2 (ICPC-2). We think that training GPs in data coding (basic and continuous training) should have a positive influence on their participation.

Moreover, most projects/networks give regular feedback reports on data recording to GPs [16, 39, 41]. In France, this feedback could represent a complementary method to assess the data required for calculating the GPs' remuneration based on public health objectives. Moreover, they could improve GPs' knowledge of their activity. Indeed, GPs are increasingly interested in data reuse at an individual level as described in the next section.

GPs role in the global network organization First, GPs are essential data providers. As care pathway coordinators [6], they represent an access point to the medical history of most of the population. GPs are by definition key players in the data collection process of primary care routine data collection networks. In France, GPs are also care pathway coordinators and therefore, this can help the development of primary care data collection projects.

Then, their role is proactive in the extraction project so as to improve data quality, that is to say reliable and reproducible data. Not only do they provide data, but they also ensure data quality. They can benefit from software training sessions and they often provide formatted or coded data to data warehouses.

Besides, GPs are also interested in data reuse at an individual level. Thanks to training, they learn to code their information and ease their abilities to extract information about their patients. GPs involved in the CPRD extraction process can receive regular practice-level quality improvement feedback, helping them improve clinical outcomes for their patients [15]. For instance, feedback on antibiotics use can be interesting so as to analyze medical practice. In the context of pay for performance, a better understanding of their patients can allow them to benefit from financial incentives.

Moreover, they can be involved at other levels of the network. For instance, department of family medicine are often members of the routine data collection networks as shown by our analysis of stakeholders in section "Stakeholders of routine primary care data collection projects". Thus, they are directly implicated in data reuse process via the selection of the data extracted and the knowledge necessary around these data. Indeed, data mining on a large scale requires a deep knowledge of the data analyzed, via the development of quality control tools. It also requires expertise to generate relevant research questions.

Lastly, due to the privacy-sensitive nature of the information, GPs represent one of the guarantees of the

integrity of routine data collection projects. This intrinsic feature facilitates the adhesion of patients and other GPs, to achieve and maintain their partnership.

Governance differences

Governance relates to "the processes of interaction and decision-making among the actors involved in a collective problem that lead to the creation, reinforcement, or reproduction of social norms and institutions." [45]. We could identify three major actors that support primary care data collection projects: private-sector companies, governmental services and academic institutions.

Private-sector companies: From extraction tools to privately funded collection projects In the absence of publicly funded software tools, such as the American VISTA software, the involvement of private-sector software companies seems to be unavoidable. We observed various levels of implications.

First, EHR software companies are involved in the development of extraction tools. Indeed, data transfer depends on the database scheme that is only known by software companies. After extraction, data are normally analyzed by academic partners, often including GPs. Indeed, very few projects were completely (data extraction and analysis) managed by private-sector companies, without the help of other official parties. For instance, IMS Health© extracts, mines and commercializes health information. In France, Cegedim© retrieves EHR data that have been coded by GPs. In such cases, the GPs' role is limited to supplying data and coding some information.

Governmental services and academic institutions The review of the main primary care data collection projects highlighted that frequently governmental services establish partnerships with academic institutions and that GPs are often part of an academic partnership. The specific role (funding, ethical assessment or data warehousing) of each party was difficult to determine.

The USA recently chose to legislate to promote data reuse; however, in our study we could not determine the effect of such legislation on the development of routine primary care data collection projects. Information was rarely available on USA data warehouses and we did not receive any answer to our enquiry by email.

Universities via their Department of General Practice are stakeholders in many projects. This involvement represents GPs' interests, and can also promote GPs' participation and be source of research project proposals, strengthening the GPs' involvement from data extraction up to data mining. Moreover, the partnership with university departments connects GPs with official institutions and with the academic community.

In this review, we identified elements that influence the development and durability of large data collection projects. It will be interesting to extend and deepen our knowledge of these databases: their purpose, content, form (structured or free-text data), quality. Indeed, in the age of Big Data, new methods of data analysis are appearing. Artificial intelligence and its deep learning methods can not only provide clinical decision support systems [46, 47] but also abilities to analyze free-text information from a large amount of data through new natural language processing algorithms.

Furthermore, the management of privacy, which is a major deterrent for GPs, is still to be addressed and should also be studied.

Conclusion

We performed a systematic review with the aim of determining the factors that allow the creation and expansion of routine data collection projects in primary care.

Technological infrastructure influence the outreach of data collection projects, particularly EHR software tools for data extraction. Beyond this first step, the most successful projects also developed comprehensive IT platforms for research purposes.

As GPs are often care pathway coordinators in most countries, primary care data are particularly important for improving healthcare management. Therefore, GPs' investment in these projects is often promoted with financial benefits, training sessions, feedback reports and involvement in research studies.

Finally, we emphasize the concomitant involvement of three main actors in supporting these initiatives: governmental services, academic institutions and software companies. Their partnership seems to be the most effective way to fund long-term and wide range data collection projects.

Several issues still need to be addressed, such as the nature of the data analyzed (coded or free-text data) or the management of privacy, which is a major deterrent for GPs. As the collected data are very sensitive, careful monitoring should be put in place due to the privacy issues at stake. Patients' de-identification should also be studied. Moreover, the project governance and its managers should be extremely transparent to improve GPs' adherence to routine primary care data collection projects.

The table does not include isolated projects that did not refer to reusable databases, and the TRANSFoRM project, which is a multinational project within Europe based on national databases, such as Nivel in the Netherlands.

Abbreviations

BHCS: Baylor Health Care System; CCHS: Christiana Care Health System; CHCI: Community Health Center, Inc.; CPCSSN: Canadian Primary Care Sentinel Surveillance Network; CPRD: Clinical Practice Research Datalink;

DELPHI: Deliver Primary Healthcare Information; EHR: Electronic Health Record; EMRALD: Electronic Medical Record Administrative Data Linked Database; FIRE: Family Medicine ICPC-Research using Electronic Medical Records; GP: General Practitioner; GRAPHC: Geographical and Resource Analysis in Primary Health Care; IFH: Institute for Family Health; IPCI: Integrated Primary Care Information project; JPCRN: Julius Primary Care Research Network; MEGPN: Melbourne East General Practice Network in Victoria project; MGPO: Massachusetts General Physicians Organization; OCHIN: Inc. Oregon Community Health Information Network; PHINEX: University of Wisconsin Electronic Health Record - Public Health Information Exchange; THIN: The Health Improvement Network; UK: United Kingdom; USA: United States of America

Acknowledgements

We thank the University of Rennes 1 librarians who assisted us in the automated literature search.

We thank E. Andermarcher for critical reading and English correction of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Authors' contributions

The work presented here was carried out in collaboration by all authors. MG, MC, AC formed the conception and study design. MG, CH and TL did the literature review, and undertook data acquisition and data analysis. MG drafted the manuscript. MC, AC, AB, LF, GB and ER supervised the review method and data interpretation and gave valuable suggestions. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of General Practice, University of Rennes 1, F-35000 Rennes, France. ²CIC (Clinical investigation center) INSERM 1414, F-35000 Rennes, France. ³INSERM, U1099, F-35000 Rennes, France. ⁴University of Rennes 1, LTSI (Laboratory for signal and image processing), F-35000 Rennes, France. ⁵CHU Rennes, CIC Inserm 1414, F-35000 Rennes, France. ⁶CHU Rennes, Centre de Données Cliniques, F-35000 Rennes, France.

Received: 23 December 2016 Accepted: 14 September 2017

Published online: 25 September 2017

References

1. Elkin PL, Trusko BE, Koppel R, Speroff T, Mohrer D, Sakji S, et al. Secondary use of clinical data. *Stud Health Technol Inform.* 2010;155:14–29.
2. Cuggia M, Garcelon N, Campillo-Gimenez B, Bernicot T, Laurent J-F, Garin E, et al. Roogle: an information retrieval engine for clinical data warehouse. *Stud Health Technol Inform.* 2011;169:584–8.
3. Metzger M-H, Durand T, Lallich S, Salamon R, Castets P. The use of regional platforms for managing electronic health records for the production of regional public health indicators in France. *BMC Med Inform Decis Mak.* 2012;12:28.
4. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: the case of the EHR4CR project. *J Biomed Inform.* 2015;53:162–73.

5. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Fam Pract*. 2006;23:253–63.
6. Masseria C, Irwin R, Thomson S, Gemmil M, Mossialos E. Primary Care in Europe [Internet]. The London school of economics and political science; Available from: <https://www.google.fr/url?sa=t&rc=jq=&esrc=s&source=web&cd=1&ved=0ahUKewIR-KWckuHMAhVCOMAKHfRYD11QFggiMAA&url=http%3A%2F%2Fec.europa.eu%2Fsocial%2FblobServlet%3FdocId%3D4739%26langId%3Den&usq=AFQjCNERLqS7id4v1IAYMuS6NffayG8uhQ&sig2=HGrl6yQOQFKi0lxbxHLOKQ&cad=rja>.
7. Avenin G. Les Bases de données issues des Dossiers Médicaux Electroniques en France. Problèmes méthodologiques et perspectives. Paris: Université Paris Descartes Faculté de Médecine Paris Descartes; 2007. [cited 2016 Jun 27]. Available from: <http://docplayer.fr/12623166-Les-bases-de-donnees-issues-des-dossiers-medicaux-electroniques-en-france-problemes-methodologiques-et-perspectives.html>
8. Van Ganse E, Letrilliart L, Borne H, Morand F, Robain M, Siegrist CA. Health problems most commonly diagnosed among young female patients during visits to general practitioners and gynecologists in France before the initiation of the human papillomavirus vaccination program. *Pharmacoepidemiol Drug Saf*. 2012;21:261–8.
9. Darmon D, Laforest L, Van Ganse E, Petrazzuoli F, van Weel C, Letrilliart L. Prescription of antibiotics and anxiolytics/hypnotics to asthmatic patients in general practice: a cross-sectional study based on French and Italian prescribing data. *BMC Fam Pract*. 2015;16(1):14. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4326444/>
10. Wallace P, Delaney B, Sullivan F. Unlocking the research potential of the GP electronic care record. *Br J Gen Pract J R Coll Gen Pract*. 2013;63:284–5.
11. Shephard E, Stapley S, Hamilton W. The use of electronic databases in primary care research. *Fam Pract*. 2011;28:352–4.
12. PRISMA [Internet]. [cited 2017 Jun 22]. Available from: <http://www.prisma-statement.org/PRISMAStatement/PRISMAStatement.aspx>
13. AMSTAR - Assessing the Methodological Quality of Systematic Reviews [Internet]. [cited 2017 Jun 22]. Available from: <https://amstar.ca/Publications.php>
14. Terry AL, Stewart M, Fortin M, Wong ST, Kennedy M, Burge F, et al. Gaps in Primary Healthcare Electronic Medical Record Research and Knowledge: Findings of a Pan-Canadian Study. *Healthc Policy*. 2014;10:46–59.
15. CPRD - Clinical Practice Research Datalink [Internet]. [cited 2016 Apr 18]. Available from: <https://www.cprd.com/intro.asp>
16. THIN - The Health Improvement Network [Internet]. [cited 2016 Apr 18]. Available from: <https://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/database>
17. Pearce C, Shearer M, Gardner K, Kelly J. A division's worth of data. *Aust Fam Physician*. 2011;40:167–70.
18. Tolar M, Balka E. Caring for individual patients and beyond: enhancing care through secondary use of data in a general practice setting. *Int. J. Med. Inf*. 2012;81:461–74.
19. Siström C, McKay NL, Weilburg JB, Atlas SJ, Ferris TG. Determinants of diagnostic imaging utilization in primary care. *Am J Manag Care*. 2012;18e135–44.
20. Wilson G, Hasnain-Wynia R, Hauser D, Calman N. Implementing Institute of Medicine recommendations on collection of patient race, ethnicity, and language data in a community health center. *J Health Care Poor Underserved*. 2013;24:875–84.
21. Parsons A, McCullough C, Wang J, Shih S. Validity of electronic health record-derived quality measurement for performance monitoring. *J Am Med Inform Assoc JAMIA*. 2012;19:604–9.
22. Devoe JE, Sears A. The OCHIN community information network: bringing together community health centers, information technology, and data to support a patient-centered medical village. *J Am Board Fam Med JABFM*. 2013;26:271–8.
23. De Clercq E, van Casteren V, Bossuyt N, Goderis G, Moreels S. Belgian primary care EPR: assessment of nationwide routine data extraction. *Stud Health Technol Inform*. 2014;197:85–9.
24. CPCSSN Data for Research - Canadian Primary Care Sentinel Surveillance Network [Internet]. [cited 2016 May 3]. Available from: <http://cpcssn.ca/research-resources/cpcssn-data-for-research/>
25. Harris SB, Glazier RH, Tompkins JW, Wilton AS, Chevendra V, Stewart MA, et al. Investigating concordance in diabetes diagnosis between primary care charts (electronic medical records) and health administrative data: a retrospective cohort study. *BMC Health Serv Res*. 2010;10:347.
26. Chmiel C, Bhend H, Senn O, Zoller M, Rosemann T. FIRE study-group. The FIRE project: a milestone for research in primary care in Switzerland. *Swiss Med Wkly*. 2011;140:w13142.
27. Melbourne East GP Network [Internet]. [cited 2016 May 13]. Available from: <https://megpn.com.au/>
28. Corporate Data Warehouse (CDW) [Internet]. [cited 2016 Apr 29]. Available from: http://www.hsrdr.research.va.gov/for_researchers/vinci/cdw.cfm
29. IPCI - Interdisciplinary Processing of Clinical Information [Internet]. [cited 2016 May 9]. Available from: <http://www.ipci.nl/Framework/Framework.php>
30. Geisinger Health System [Internet]. [cited 2016 May 2]. Available from: <https://www.geisinger.org/>
31. SIDiAP - Information System for the Improvement of Research in Primary Care [Internet]. [cited 2016 May 2]. Available from: <http://www.sidiap.org/>
32. Phillips RL, Bazemore AW, DeVoe JE, Weida TJ, Krist AH, Dulin MF, et al. A Family Medicine Health Technology Strategy for Achieving the Triple Aim for US Health Care. *Fam Med*. 2015;47:628–35.
33. Yaeger JP, Temte JL, Hanrahan LP, Martinez-Donate P. Roles of Clinician, Patient, and Community Characteristics in the Management of Pediatric Upper Respiratory Tract Infections. *Ann Fam Med*. 2015;13:529–36.
34. Tu K, Mitiku TF, Ivers NM, Guo H, Lu H, Jaakkimainen L, et al. Evaluation of Electronic Medical Record Administrative data Linked Database (EMRALD). *Am J Manag Care*. 2014;20:e15–21.
35. Widdifield J, Bombardier C, Bernatsky S, Paterson JM, Green D, Young J, et al. An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. *BMC Musculoskelet Disord*. 2014;15:216.
36. Greiver M, Williamson T, Bennett T-L, Drummond N, Savage C, Aliarzadeh B, et al. Developing a method to estimate practice denominators for a national Canadian electronic medical record database. *Fam Pract*. 2013;30:347–54.
37. Tu K, Wang M, Jaakkimainen RL, Butt D, Ivers NM, Young J, et al. Assessing the validity of using administrative data to identify patients with epilepsy. *Epilepsia*. 2014;55:335–43.
38. Ferrajolo C, Verhamme KMC, Trifirò G, t Jong GW, Giaquinto C, Picelli G, et al. Idiopathic acute liver injury in paediatric outpatients: incidence and signal detection in two European countries. *Drug Saf*. 2013;36:1007–16.
39. JPCRN - Julius primary care research network [Internet]. [cited 2016 May 13]. Available from: <https://primarycare.juliusclinical.com/pcrn/?lang=en>
40. Quint JK, Müllerova H, DiSantostefano RL, Forbes H, Eaton S, Hurst JR, et al. Validation of chronic obstructive pulmonary disease recording in the Clinical Practice Research Datalink (CPRD-GOLD). *BMJ Open*. 2014;4:e005540.
41. NIVEL - Netherlands institute for health services research [Internet]. [cited 2016 May 9]. Available from: <https://www.nivel.nl/en>
42. QResearch - What Is QResearch [Internet]. [cited 2016 Apr 29]. Available from: <http://www.qresearch.org/SitePages/What%20Is%20QResearch.aspx>
43. Shin P, Sharac J. Readiness for meaningful use of health information technology and patient centered medical home recognition survey results. *Medicare Medicaid Res Rev*. 2013;3:4.
44. Darmon D, Sauvart R, Staccini P, Letrilliart L. Which functionalities are available in the electronic health record systems used by French general practitioners? An assessment study of 15 systems. *Int. J. Med. Inf*. 2014;83:37–46.
45. Hufty M. Investigating Policy Processes: The Governance Analytical Framework (GAF) [Internet], Report No.: ID 2019005. Rochester: Social Science Research Network; 2011. p. 403–24. Available from: <http://papers.ssrn.com/abstract=2019005>
46. Kruskal JB, Berkowitz S, Geis JR, Kim W, Nagy P, Dreyer K. Big Data and Machine Learning-Strategies for Driving This Bus: A Summary of the 2016 Intersociety Summer Conference. *J Am Coll Radiol JACR*. 2017;14:811–7.
47. Tripoliti EE, Papadopoulos TG, Karanasiou GS, Naka KK, Fotiadis DI. Heart Failure: Diagnosis, Severity Estimation and Prediction of Adverse Events Through Machine Learning Techniques. *Comput Struct Biotechnol J*. 2017;15:26–47.
48. Mazumdar S, Konings P, Hewett M, Bagheri N, McRae I, Del Fante P. Protecting the privacy of individual general practice patient electronic records for geospatial epidemiology research. *Aust N Z J Public Health*. 2014;38:548–52.
49. GRAPHIC. Geographic Resource and Analysis in Primary Health Care. Aust Prim Health Care Res Inst. 2014; [cited 2016 May 16]. Available from: <http://aphcri.anu.edu.au/graphic>
50. Liljeqvist GTH, Staff M, Puech M, Blom H, Torvaldsen S. Automated data extraction from general practice records in an Australian setting: trends in influenza-like illness in sentinel general practices and emergency departments. *BMC Public Health*. 2011;11:435.

51. Young J, Eley D, Fahey P, Patterson E, Hegney D. Enabling research in general practice—increasing functionality of electronic medical records. *Aust Fam Physician*. 2010;39:506–9.
52. Truyers C, Goderis G, Dewitte H, Akker M, van den Buntinx F. The Intego database: background, methods and basic results of a Flemish general practice-based continuous morbidity registration project. *BMC Med Inform Decis Mak*. 2014;14:48.
53. Bossuyt N, Van Casteren V, Goderis G, Wens J, Moreels S, Vanthomme K, et al. Public Health Triangulation to inform decision-making in Belgium. *Stud Health Technol Inform*. 2015;210:855–9.
54. Intego [Internet]. [cited 2016 May 13]. Available from: <https://intego.be/en/Welcome>
55. Coleman N, Halas G, Peeler W, Casalang N, Williamson T, Katz A. From patient care to research: a validation study examining the factors contributing to data quality in a primary care electronic medical record database. *BMC Fam Pract*. 2015;16:11.
56. Williamson T, Green ME, Birtwhistle R, Khan S, Garies S, Wong ST, et al. Validating the 8 CPCSSN Case Definitions for Chronic Disease Surveillance in a Primary Care Database of Electronic Health Records. *Ann Fam Med*. 2014;12:367–72.
57. Williamson T, Lévesque L, Morkem R, Birtwhistle R. CPCSSN's role in improving pharmacovigilance. *Can Fam Physician Médecin Fam Can*. 2014;60:678–80.
58. Keshavjee K, Williamson T, Martin K, Truant R, Aliazadeh B, Ghany A, et al. Getting to usable EMR data. *Can Fam Physician Médecin Fam Can*. 2014;60:392.
59. Torti J, Duerksen K, Forst B, Salvalaggio G, Jackson D, Manca D. Documenting alcohol use in primary care in Alberta. *Can Fam Physician Médecin Fam Can*. 2013;59(1128):e473–4.
60. Garies S, Jackson D, Aliazadeh B, Keshavjee K, Martin K, Williamson T. Sentinel eye: improving usability of smoking data in EMR systems. *Can Fam Physician Médecin Fam Can*. 2013;59(108):e60–1.
61. Greiver M, Keshavjee K, Jackson D, Forst B, Martin K, Aliazadeh B. Sentinel feedback: path to meaningful use of EMRs. *Can Fam Physician Médecin Fam Can*. 2012;58(1168):e611–2.
62. Birtwhistle R, Williamson T. Primary care electronic medical records: a new data source for research in Canada. *CMAJ Can Med Assoc J J Assoc Medicale Can*. 2015;187:239–40.
63. Barber D, Williamson T, Biro S, Hall Barber K, Martin D, Kinsella L, et al. Data discipline in electronic medical records: Improving smoking status documentation with a standardized intake tool and process. *Can Fam Physician Med Fam Can*. 2015;61:e570–6.
64. Morkem R, Barber D, Williamson T, Patten SB. A Canadian Primary Care Sentinel Surveillance Network Study Evaluating Antidepressant Prescribing in Canada From 2006 to 2012. *Can J Psychiatr Rev Can Psychiatr*. 2015;60:564–70.
65. Farahani P, Khan S, Oatway M, Dziarmaga A. Exploring the Distribution of Prescription for Sulfonylureas in Patients with Type 2 Diabetes According to Cardiovascular Risk Factors Within a Canadian Primary Care Setting. *J Popul Ther Clin Pharmacol J Ther Popul Pharmacologie Clin*. 2015;22:e228–36.
66. Greiver M, Aliazadeh B, Meaney C, Moineddin R, Southgate CA, Barber DTS, et al. Are We Asking Patients if They Smoke?: Missing Information on Tobacco Use in Canadian Electronic Medical Records. *Am J Prev Med*. 2015;49:264–8.
67. Maddocks H, Marshall JN, Stewart M, Terry AL, Cejic S, Hammond J-A, et al. Quality of congestive heart failure care: assessing measurement of care using electronic medical records. *Can Fam Physician Médecin Fam Can*. 2010;56:e432–7.
68. Allin S, Munce S, Jaglal S, Butt D, Young J, Tu K. Capture of osteoporosis and fracture information in an electronic medical record database from primary care. *AMIA Annu Symp Proc AMIA Symp AMIA Symp*. 2014;2014:240–8.
69. Tu K, Wang M, Young J, Green D, Ivers NM, Butt D, et al. Validity of administrative data for identifying patients who have had a stroke or transient ischemic attack using EMERALD as a reference standard. *Can J Cardiol*. 2013;29:1388–94.
70. Schultz SE, Rothwell DM, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. *Chronic Dis Inj Can*. 2013;33:160–6.
71. Butt DA, Tu K, Young J, Green D, Wang M, Ivers N, et al. A validation study of administrative data algorithms to identify patients with Parkinsonism with prevalence and incidence trends. *Neuroepidemiology*. 2014;43:28–37.
72. Tu JV, Chu A, Donovan LR, Ko DT, Booth GL, Tu K, et al. The Cardiovascular Health in Ambulatory Care Research Team (CANHEART): using big data to measure and improve cardiovascular health and healthcare services. *Circ Cardiovasc Qual Outcomes*. 2015;8:204–12.
73. Shadd JD, Ryan BL, Maddocks HL, McKay SD, Moulin DE. Neuropathic pain in a primary care electronic health record database. *Eur J Pain Lond Engl*. 2015;19:715–21.
74. EMERALD - Electronic Medical Record Administrative Data Linked Database [Internet]. [cited 2016 Apr 25]. Available from: <http://www.ices.on.ca/Research/Research-programs/Primary-Care-and-Population-Health/EMERALD>
75. Cricelli I, Lapi F, Montalbano C, Medea G, Cricelli C. Mille general practice governance (MilleGPG): an interactive tool to address an effective quality of care through the Italian general practice network. *Prim Health Care Res Dev*. 2013;14:409–12.
76. Valkhoff VE, Coloma PM, Masdee GMC, Gini R, Innocenti F, Lapi F, et al. Validation study in four health-care databases: upper gastrointestinal bleeding misclassification affects precision but not magnitude of drug-related upper gastrointestinal bleeding risk. *J Clin Epidemiol*. 2014;67:921–31.
77. Health search database - Italy [Internet]. [cited 2016 May 13]. Available from: <https://www.healthsearch.it/?lang=en>
78. Soler JK, Corrigan D, Kazienko P, Kajdanowicz T, Danger R, Kulisiewicz M, et al. Evidence-based rules from family practice to inform family practice; the learning healthcare system case study on urinary tract infections. *BMC Fam Pract*. 2015;16:63.
79. TRANSFoRm project [Internet]. [cited 2016 May 16]. Available from: <http://www.transformproject.eu/>
80. van Lier A, van Erp J, Donker GA, van der Maas NAT, Sturkenboom MCJM, de Melker HE. Low varicella-related consultation rate in the Netherlands in primary care data. *Vaccine*. 2014;32:3517–24.
81. Afzal Z, Engelkes M, Verhamme KMC, Janssens HM, Sturkenboom MCJM, Kors JA, et al. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. *Pharmacoepidemiol Drug Saf*. 2013;22:826–33.
82. van Wyk JT, Mosseveld B, van der Lei J. Is population-oriented IT supported preventive care in general practice feasible? A database study. *Stud Health Technol Inform*. 2010;160:462–5.
83. Woestenberg PJ, van Oeffelen AAM, Stirbu-Wagner I, van Benthem BHB, van Bergen JEAM, van den Broek IVF. Comparison of STI-related consultations among ethnic groups in the Netherlands: an epidemiologic study using electronic records from general practices. *BMC Fam Pract*. 2015;16:70.
84. Kuchinke W, Ohmann C, Verheij RA, van Veen E-B, Arvanitis TN, Taweel A, et al. A standardised graphic method for describing data privacy frameworks in primary care research using a flexible zone model. *Int J Med Inf*. 2014;83:941–57.
85. Nielsen MMJ, Ursum J, Schellevis FG, Korevaar JC. The validity of the diagnosis of inflammatory arthritis in a large population-based primary care database. *BMC Fam Pract*. 2013;14:79.
86. Opondo D, Visscher S, Eslami S, Verheij RA, Korevaar JC, Abu-Hanna A. Quality of Co-Prescribing NSAID and Gastroprotective Medications for Elders in The Netherlands and Its Association with the Electronic Medical Record. *PLoS One*. 2015;10:e0129515.
87. Dentler K, Numans ME, ten Teije A, Cornet R, de Keizer NF. Formalization and computation of quality measures based on electronic medical records. *J Am Med Inform Assoc JAMIA*. 2014;21:285–91.
88. Smits FT, Brouwer HJ, Zwinderman AH, van den Akker M, van Steenkiste B, Mohrs J, et al. Predictability of persistent frequent attendance in primary care: a temporal and geographical validation study. *PLoS One*. 2013;8:e73125.
89. Registration Network Family Practices - Maastricht - Netherlands [Internet]. [cited 2016 Jul 18]. Available from: <http://www.generalpracticemaastricht.nl/what-we-do/networks/>
90. Garcia-Gil M, Elorza J-M, Banque M, Comas-Cufí M, Blanch J, Ramos R, 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:e109706.
91. Garcia-Gil MDM, Hermsilla E, Prieto-Alhambra D, Fina F, Rosell M, Ramos R, 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;19:135–45.
92. Bolibar B, Fina Avilés F, Morros R, Garcia-Gil M del M, Hermsilla E, Ramos R, et al. SIDIAP database: electronic clinical records in primary care as a source of information for epidemiologic research. *Med Clínica*. 2012;138:617–21.
93. Esteban-Vasallo MD, Gil-Prieto R, Domínguez-Berjón MF, Astray-Mochales J, Gil de Miguel A. Temporal trends in incidence rates of herpes zoster among patients treated in primary care centers in Madrid (Spain), 2005-2012. *J Inf Secur*. 2014;68:378–86.

94. Gil M, Oliva B, Timoner J, Maciá 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;72:965–8.
95. BIFAP - Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria [Internet]. [cited 2016 May 13]. Available from: <http://www.bifap.org/summary.php>
96. Neumark T, Brudin L, Mölstad S. Use of rapid diagnostic tests and choice of antibiotics in respiratory tract infections in primary healthcare—a 6-y follow-up study. *Scand J Infect Dis*. 2010;42:90–6.
97. Streit S, Kaplan V, Busato A, Djalali S, Senn O, Meli DN, et al. General Practitioners' vitamin K antagonist monitoring is associated with better blood pressure control in patients with hypertension—a cross-sectional database study. *BMC Cardiovasc Disord*. 2015;15:47.
98. FIRE project [Internet]. [cited 2016 May 13]. Available from: <http://www.hausarztmedizin.uzh.ch/de.html>
99. Springate DA, Ashcroft DM, Kontopantelis E, Doran T, Ryan R, Reeves D. Can analyses of electronic patient records be independently and externally validated? Study 2—the effect of β -adrenoceptor blocker therapy on cancer survival: a retrospective cohort study. *BMJ Open*. 2015;5:e007299.
100. Cornish R, Tilling K, Boyd A, Macleod J, Van Staa T. Using linkage to electronic primary care records to evaluate recruitment and nonresponse bias in the Avon Longitudinal Study of Parents and Children. *Epidemiol Camb Mass*. 2015;26:e41–2.
101. Sammon CJ, Miller A, Mahtani KR, Holt TA, McHugh NJ, Luqmani RA, et al. Missing laboratory test data in electronic general practice records: analysis of rheumatoid factor recording in the clinical practice research datalink. *Pharmacoepidemiol Drug Saf*. 2015;24:504–9.
102. Wurst KE, Shukla A, Muellerova H, Davis KJ. Respiratory pharmacotherapy use in patients newly diagnosed with chronic obstructive pulmonary disease in a primary care setting in the UK: a retrospective cohort study. *COPD*. 2014;11:521–30.
103. Dregan A, Charlton J, Wolfe CDA, Gulliford MC, Markus HS. Is sodium valproate, an HDAC inhibitor, associated with reduced risk of stroke and myocardial infarction? A nested case-control study. *Pharmacoepidemiol Drug Saf*. 2014;23:759–67.
104. Cornish RP, Henderson J, Boyd AW, Granell R, Van Staa T, Macleod J. Validating childhood asthma in an epidemiological study using linked electronic patient records. *BMJ Open*. 2014;4:e005345.
105. Reeves D, Springate DA, Ashcroft DM, Ryan R, Doran T, Morris R, 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:e004952.
106. Muller S. Electronic medical records: the way forward for primary care research? *Fam Pract*. 2014;31:127–9.
107. Rushton CA, Strömberg A, Jaarsma T, Kadam UT. Multidrug and optimal heart failure therapy prescribing in older general practice populations: a clinical data linkage study. *BMJ Open*. 2014;4:e003698.
108. Tate AR, Beloff N, Al-Radwan B, Wickson J, Puri S, Williams T, et al. Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and an intuitive query interface. *J Am Med Assoc*. 2014;312:292–8.
109. Hammad TA, Margulis AV, Ding Y, Strazzeri MM, Epperly H. Determining the predictive value of Read codes to identify congenital cardiac malformations in the UK Clinical Practice Research Datalink. *Pharmacoepidemiol Drug Saf*. 2013;22:1233–8.
110. Ford E, Nicholson A, Koeling R, Tate A, Carroll J, Axelrod L, 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.
111. Denaxas SC, George J, Herrett E, Shah AD, Kalra D, Hingorani AD, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol*. 2012;41:1625–38.
112. Dregan A, van Staa T, McDermott L, McCann G, Ashworth M, Charlton J, et al. Cluster randomized trial in the general practice research database: 2. Secondary prevention after first stroke (eCRT study): study protocol for a randomized controlled trial. *Trials*. 2012;13:181.
113. Dregan A, Moller H, Murray-Thomas T, Gulliford MC. Validity of cancer diagnosis in a primary care database compared with linked cancer registrations in England. Population-based cohort study. *Cancer Epidemiol*. 2012;36:425–9.
114. Taylor A, Stapley S, Hamilton W. Jaundice in primary care: a cohort study of adults aged >45 years using electronic medical records. *Fam Pract*. 2012;29:416–20.
115. Chen Y-C, Wu J-C, Haschler I, Majeed A, Chen T-J, Wetter T. Academic impact of a public electronic health database: bibliometric analysis of studies using the general practice research database. *PLoS One*. 2011;6:e21404.
116. Herrett EL, Thomas SL, Smeeth L. Validity of diagnoses in the general practice research database. *Br J Gen Pract J R Coll Gen Pract*. 2011;61:438–9.
117. Dregan A, Toschke MA, Wolfe CD, Rudd A, Ashworth M, Gulliford MC, et al. Utility of electronic patient records in primary care for stroke secondary prevention trials. *BMC Public Health*. 2011;11:86.
118. Nicholson A, Rait G, Murray-Thomas T, Hughes G, Mercer CH, Cassell J. Management of epididymo-orchitis in primary care: results from a large UK primary care database. *Br J Gen Pract J R Coll Gen Pract*. 2010;60:e407–22.
119. Nicholson A, Rait G, Murray-Thomas T, Hughes G, Mercer CH, Cassell J. Management of first-episode pelvic inflammatory disease in primary care: results from a large UK primary care database. *Br J Gen Pract J R Coll Gen Pract*. 2010;60:e395–406.
120. Kurd SK, Troxel AB, Crits-Christoph P, Gelfand JM. The risk of depression, anxiety, and suicidality in patients with psoriasis: a population-based cohort study. *Arch Dermatol*. 2010;146:891–5.
121. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract J R Coll Gen Pract*. 2010;60:e128–36.
122. Johansson S, Wallander M-A, de Abajo FJ, García Rodríguez LA. Prospective drug safety monitoring using the UK primary-care General Practice Research Database: theoretical framework, feasibility analysis and extrapolation to future scenarios. *Drug Saf*. 2010;33:223–32.
123. Devine S, West S, Andrews E, Tennis P, Hammad TA, Eaton S, et al. The identification of pregnancies within the general practice research database. *Pharmacoepidemiol Drug Saf*. 2010;19:45–50.
124. Olier I, Springate DA, Ashcroft DM, Doran T, Reeves D, Planner C, et al. Modelling Conditions and Health Care Processes in Electronic Health Records: An Application to Severe Mental Illness with the Clinical Practice Research Datalink. *PLoS One*. 2016;11:e0146715.
125. Gallagher AM, Williams T, Leufkens HGM, de Vries F. The Impact of the Choice of Data Source in Record Linkage Studies Estimating Mortality in Venous Thromboembolism. *PLoS One*. 2016;11:e0148349.
126. Pujades-Rodriguez M, Duyx B, Thomas SL, Stogiannidis D, Smeeth L, Hemingway H. Associations between polymyalgia rheumatica and giant cell arteritis and 12 cardiovascular diseases. *Heart Br Card Soc*. 2016;102:383–9.
127. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44:827–36.
128. Stevenson F. The use of electronic patient records for medical research: conflicts and contradictions. *BMC Health Serv Res*. 2015;15:124.
129. Charlton RA, Neville AJ, Jordan S, Pierini A, Damase-Michel C, Klungsoyr K, et al. Healthcare databases in Europe for studying medicine use and safety during pregnancy. *Pharmacoepidemiol Drug Saf*. 2014;23:586–94.
130. 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.
131. Irizarry MC, Webb DJ, Boudiaf N, Logie J, Habel LA, Udaltsova N, et al. Risk of cancer in patients exposed to gabapentin in two electronic medical record systems. *Pharmacoepidemiol Drug Saf*. 2012;21:214–25.
132. van Staa TP, Patel D, Gallagher AM, de Bruin ML. Glucose-lowering agents and the patterns of risk for cancer: a study with the General Practice Research Database and secondary care data. *Diabetologia*. 2012;55:654–65.
133. Miller DP, Watkins SE, Sampson T, Davis KJ. Long-term use of fluticasone propionate/salmeterol fixed-dose combination and incidence of cataracts and glaucoma among chronic obstructive pulmonary disease patients in the UK General Practice Research Database. *Int J Chron Obstruct Pulmon Dis*. 2011;6:467–76.
134. Walker AM. Identification of esophageal cancer in the General Practice Research Database. *Pharmacoepidemiol Drug Saf*. 2011;20:1159–67.
135. Mamtani R, Haynes K, Finkelman BS, Scott FI, Lewis JD. Distinguishing incident and prevalent diabetes in an electronic medical records database. *Pharmacoepidemiol Drug Saf*. 2014;23:111–8.
136. Hall GC, Hill F. Descriptive investigation of the recording of influenza vaccination details on The Health Information Network database. *Pharmacoepidemiol Drug Saf*. 2014;23:595–600.

137. Horsfall L, Walters K, Petersen I. Identifying periods of acceptable computer usage in primary care research databases. *Pharmacoepidemiol Drug Saf.* 2013;22:64–9.
138. Haynes K, Bilker WB, Tenhave TR, Strom BL, Lewis JD. Temporal and within practice variability in the health improvement network. *Pharmacoepidemiol Drug Saf.* 2011;20:948–55.
139. Toh S, García Rodríguez LA, Hernán MA. Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records. *Pharmacoepidemiol Drug Saf.* 2011;20:849–57.
140. Staff M. Can data extraction from general practitioners' electronic records be used to predict clinical outcomes for patients with type 2 diabetes? *Inform Prim Care.* 2012;20:95–102.
141. Langley TE, Huang Y, Lewis S, McNeill A, Coleman T, Szatkowski L. Prescribing of nicotine replacement therapy to adolescents in England. *Addiction.* 2011;106:1513–9.
142. Szatkowski L, Lewis S, McNeill A, Coleman T. Is smoking status routinely recorded when patients register with a new GP? *Fam Pract.* 2010;27:673–5.
143. Ruigómez A, Martín-Merino E, Rodríguez LAG. Validation of ischemic cerebrovascular diagnoses in the health improvement network (THIN). *Pharmacoepidemiol Drug Saf.* 2010;19:579–85.
144. Meropol SB, Metlay JP. Accuracy of pneumonia hospital admissions in a primary care electronic medical record database. *Pharmacoepidemiol Drug Saf.* 2012;21:659–65.
145. White D, Choi H, Peloquin C, Zhu Y, Zhang Y. Secular trend of adhesive capsulitis. *Arthritis Care Res.* 2011;63:1571–5.
146. Kalankesh L, Weatherall J, Ba-Dhfari T, Buchan I, Brass A. Taming EHR data: using semantic similarity to reduce dimensionality. *Stud Health Technol Inform.* 2013;192:52–6.
147. Torjesen I. £40m is wasted on GP data extraction IT system that does not work properly. *BMJ.* 2015;351:h3609.
148. Mason B, Boyd K, Murray SA, Steyn J, Cormie P, Kendall M, et al. Developing a computerised search to help UK General Practices identify more patients for palliative care planning: a feasibility study. *BMC Fam Pract.* 2015;16:99.
149. Mohamed IN, Helms PJ, Simpson CR, Milne RM, McLay JS. Using primary care prescribing databases for pharmacovigilance. *Br J Clin Pharmacol.* 2011;71:244–9.
150. Poh N, McGovern AP, de Lusignan S. Improving the measurement of longitudinal change in renal function: automated detection of changes in laboratory creatinine assay. *J Innov Health Inform.* 2015;22:293–301.
151. Elkhenini HF, Davis KJ, Stein ND, New JP, Delderfield MR, Gibson M, et al. Using an electronic medical record (EMR) to conduct clinical trials: Salford Lung Study feasibility. *BMC Med Inform Decis Mak.* 2015;15:8.
152. Murtaugh MA, Gibson BS, Redd D, Zeng-Treitler Q. Regular expression-based learning to extract bodyweight values from clinical notes. *J Biomed Inform.* 2015;54:186–90.
153. Lisi AJ, Burgo-Black AL, Kawecki T, Brandt CA, Goulet JL. Use of Department of Veterans Affairs administrative data to identify veterans with acute low back pain: a pilot study. *Spine.* 2014;39:1151–6.
154. Redd D, Kuang J, Zeng-Treitler Q. Differences in nationwide cohorts of acupuncture users identified using structured and free text medical records. *AMIA Annu Symp Proc AMIA Symp AMIA Symp.* 2014;2014:1002–9.
155. Wang L, Porter B, Maynard C, Evans G, Bryson C, Sun H, et al. Predicting risk of hospitalization or death among patients receiving primary care in the Veterans Health Administration. *Med Care.* 2013;51:368–73.
156. McCart JA, Berndt DJ, Jarman J, Finch DK, Luther SL. Finding falls in ambulatory care clinical documents using statistical text mining. *J Am Med Inform Assoc JAMIA.* 2013;20:906–14.
157. Serrano N, Molander R, Monden K, Grosshans A, Krahn DD. Exemplars in the use of technology for management of depression in primary care. *WMJ Off Publ State Med Soc Wis.* 2012;111:112–8.
158. Fisher DA, Grubber JM, Castor JM, Coffman CJ. Ascertainment of colonoscopy indication using administrative data. *Dig Dis Sci.* 2010;55:1721–5.
159. Salem RM, Pandey B, Richard E, Fung MM, García EP, Brophy VH, et al. The VA Hypertension Primary Care Longitudinal Cohort: Electronic medical records in the post-genomic era. *Health Informatics J.* 2010;16:274–86.
160. Vijaykrishnan R, Steinhubl SR, Ng K, Sun J, Byrd RJ, Daar Z, et al. Prevalence of Heart Failure Signs and Symptoms in a Large Primary Care Population Identified Through the Use of Text and Data Mining of the Electronic Health Record. *J Card Fail.* 2014;20:459–64.
161. Byrd RJ, Steinhubl SR, Sun J, Ebadollahi S, Stewart WF. Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records. *Int J Med Inf.* 2014;83:983–92.
162. Wang Y, Ng K, Byrd RJ, Hu J, Ebadollahi S, Daar Z, et al. Early detection of heart failure with varying prediction windows by structured and unstructured data in electronic health records. *Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf.* 2015;2015:2530–3.
163. Concurrent control of blood glucose, body mass, and blood pressure in patients with type 2 diabetes: an analysis of data from electronic medical re... - PubMed - NCBI [Internet]. [cited 2016 May 2]. Available from: <http://www.ncbi.nlm.nih.gov/pas-serelle.univ-rennes1.fr/pubmed/21397777>
164. Crawford AG, Cote C, Couto J, Daskiran M, Gunnarsson C, Haas K, et al. Comparison of GE Centricity Electronic Medical Record database and National Ambulatory Medical Care Survey findings on the prevalence of major conditions in the United States. *Popul Health Manag.* 2010;13:139–50.
165. OCHIN [Internet]. [cited 2016 Nov 4]. Available from: <https://ochin.org/>
166. Brouwer ES, West SL, Kluckman M, Wallace D, Masica AL, Ewen E, et al. Initial and subsequent therapy for newly diagnosed type 2 diabetes patients treated in primary care using data from a vendor-based electronic health record. *Pharmacoepidemiol Drug Saf.* 2012;21:920–8.
167. Baylor health care system [Internet]. [cited 2016 Nov 7]. Available from: <http://www.baylorhealth.com/SpecialtiesServices/PrimaryCare/Pages/Default.aspx>
168. Geraghty EM, Balsbaugh T, Nuovo J, Tandon S. Using Geographic Information Systems (GIS) to assess outcome disparities in patients with type 2 diabetes and hyperlipidemia. *J Am Board Fam Med JABFM.* 2010;23:88–96.
169. UC Davis Health System [Internet]. [cited 2016 Nov 7]. Available from: <https://secure.ucdmc.ucdavis.edu/welcome/index.html>
170. Tian TY, Zlateva I, Anderson DR. Using electronic health records data to identify patients with chronic pain in a primary care setting. *J Am Med Inform Assoc JAMIA.* 2013;20:e275–80.
171. CHCI [Internet]. [cited 2016 Nov 4]. Available from: <http://www.chc1.com/>
172. Massachusetts General Physicians Organization [Internet]. [cited 2016 Nov 7]. Available from: <http://www.massgeneral.org/mgpo/>
173. Shelley D, Tseng T-Y, Matthews AG, Wu D, Ferrari P, Cohen A, et al. Technology-driven intervention to improve hypertension outcomes in community health centers. *Am J Manag Care.* 2011;17:SP103–10.
174. Open Door Family Medical Centers [Internet]. [cited 2016 Nov 4]. Available from: <https://www.opendoormedical.org/>
175. PCIP [Internet]. [cited 2016 Nov 7]. Available from: <http://www1.nyc.gov/site/doh/providers/electronic-records.page>
176. Institute for Family Health [Internet]. [cited 2016 Nov 13]. Available from: <http://www.institute.org/>
177. Makam AN, Nguyen OK, Moore B, Ma Y, Amarasingham R. Identifying patients with diabetes and the earliest date of diagnosis in real time: an electronic health record case-finding algorithm. *BMC Med Inform Decis Mak.* 2013;13:81.
178. Freund J, Meiman J, Kraus C. Using electronic medical record data to characterize the level of medication use by age-groups in a network of primary care clinics. *J Prim Care Community Health.* 2013;4:286–93.
179. Meiman J, Freund JE. Large data sets in primary care research. *Ann Fam Med.* 2012;10:473–4.
180. IMS Health: Real-World Data; 2016. [cited 2016 Jun 30]. Available from: <http://www.imshealth.com/en/solution-areas/real-world-evidence/rwe-solutions/real-world-data>
181. Centricity EMR - General Electric Healthcare IT [Internet]. [cited 2016 May 16]. Available from: http://www3.gehealthcare.com/en/products/categories/healthcare_it/electronic_medical_records/centricity_emr
182. Australian Primary Health Care Research Institute [Internet]. [cited 2016 May 13]. Available from: <http://aphcri.anu.edu.au/>