

**Research article**

**Outputs and growth of primary care databases in the United Kingdom: bibliometric analysis**

Cite this article: Mannan F, Chaudhry Z, Gibson-White A, Syed U, Ahmed S, Kousoulis A, Majeed A. Outputs and growth of primary care databases in the United Kingdom: bibliometric analysis. *J Innov Health Inform.* 2017;24(3):284–290.

<http://dx.doi.org/10.14236/jhi.v24i3.942>

Copyright © 2017 The Author(s). Published by BCS, The Chartered Institute for IT under Creative Commons license <http://creativecommons.org/licenses/by/4.0/>

**Author address for correspondence:**  
Zain Chaudhry  
Department of Primary Care and Public Health  
Imperial College London  
London W6 8RP, UK  
Email: [zain.chaudhry10@imperial.ac.uk](mailto:zain.chaudhry10@imperial.ac.uk)

Accepted October 2017

**Fahmida Mannan**

Department of Primary Care and Public Health, Imperial College, London, UK

**Zain Chaudhry**

Department of Primary Care and Public Health, Imperial College, London, UK

**Angela Gibson-White**

Department of Primary Care and Public Health, Imperial College, London, UK

**Usama Syed**

Department of Primary Care and Public Health, Imperial College, London, UK

**Shirin Ahmed**

Kings College London, London, UK

**Antonios Kousoulis**

Medicines and Healthcare Products Regulatory Agency, Clinical Practice Research Datalink, London, UK

**Azeem Majeed**

Department of Primary Care and Public Health, Imperial College, London, UK

**ABSTRACT**

**Background** Electronic health database (EHD) data are increasingly used by researchers. The major United Kingdom EHDs are the ‘Clinical Practice Research Datalink’ (CPRD), ‘The Health Improvement Network’ (THIN) and ‘QResearch’. Over time, outputs from these databases have increased but have not been evaluated.

**Objective** This study compares research outputs from CPRD, THIN and QResearch assessing growth and publication outputs over a 10-year period (2004–2013). CPRD was also reviewed separately over 20 years as a case study.

**Methods** Publications from CPRD and QResearch were extracted using the Science Citation Index of the Thomson Scientific Institute for Scientific Information (Web of Science). THIN data were obtained from University College London and validated in the Web of Science. All databases were analysed for growth in publications, the speciality areas and the journals in which their data have been published.

**Results** These databases collectively produced 1,296 publications over a ten-year period, with CPRD representing 63.6% (n = 825 papers), THIN 30.4% (n = 394) and QResearch 5.9% (n = 77). Pharmacoepidemiology and General Medicine were the most common specialities featured. Over the 9-year period (2004–2013), publications for THIN and QResearch have slowly increased over time, whereas CPRD publications have increased substantially in the last 4 years with almost 75% of CPRD publications published in the past 9 years.

**Conclusion** These databases are enhancing scientific research and are growing yearly, however display variability in their growth. They could become more powerful research tools if the National Health Service and general practitioners can provide accurate and comprehensive data for inclusion in these databases.

## INTRODUCTION

Data collected in electronic medical records for a patient in primary care can span from birth to death and can have enormous benefits in improving health care and public health, and for research. Several systems exist in the United Kingdom (UK) to facilitate the use of research data generated from consultations between primary care professionals and their patients. General Practitioners play a gatekeeper role in the UK's National Health Service (NHS) because they are responsible for providing primary care services and for referring patients to see specialists.

In more recent years, these databases have been supplemented (through data linkage) with additional data from areas such as laboratory investigations, hospital admissions and mortality statistics. Data collected in primary care research databases are now increasingly used for research in many areas, and for providing information on patterns of disease.<sup>1</sup> These databases have clinical and prescription data and can provide information to support pharmacovigilance, including information on demographics, medical symptoms, therapy (medicines, vaccines, devices) and treatment outcomes.<sup>1</sup> The major primary care research databases in the UK include the 'Clinical Practice Research Datalink' (CPRD), 'QResearch' and 'The Health Improvement Network' (THIN). For all three systems, the information relating to symptoms, diseases, consultations and other clinical events is recorded using the Read code system. The data made available to researchers are anonymised, and strong patient identifiers such as name, address and postcode, date of birth and NHS number are removed.

The CPRD is jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicine and Healthcare Products Regulatory Agency (MHRA).<sup>2</sup> It is one of the largest databases of longitudinal medical records derived from primary care in the world.<sup>3</sup> The collection of information began in 1987 under the previous name General Practice Research Database (GPRD). GPRD was initially part of the Value Added Medical Products, a company that pioneered the design and marketing of a general practice office computer system, allowing the recording of individual patient medical recording. The database was later transferred to government control.<sup>4</sup> CPRD has been providing nearly 30 years of longitudinal data. As of December 2014, the database contained data for over 13.5 million patients, of which approximately 5.7 million are currently active.<sup>4</sup> As well as primary care data, CPRD now links to a number of other data sets such as 'Hospital Episode Statistics' (HES) and mortality data from the Office for National Statistics. It is increasingly being used to enhance clinical trial efficiency (protocol optimization, feasibility and recruitment), through working with the general practitioners, and can provide data for both industry and academic researchers.<sup>4</sup> Access to the data is subject to protocol approval by the MHRA Independent Scientific Advisory Committee. Over 1,500 research reports published in peer-reviewed journals have used data from the CPRD and have had direct impacts on public health and disease speciality areas.<sup>5</sup>

**QResearch** is a large primary care database derived from the anonymised health records of over 12 million patients.<sup>6</sup> The data currently come from over 950 general practices using the Egton Medical Information Systems (EMIS) clinical computer system that is used throughout the UK.<sup>6</sup> Although the data contain socio-economic details of patients based on their postcode, it does not hold any identifiable data, and access to it is only opened to academic researchers who have ethical approval to receive datasets. QResearch has led many projects such as QFlu, which was used for monitoring and tracking the prevalence of the swine flu outbreak in 2009, reporting to the Health Protection Agency.<sup>6</sup> One of the limitations of QResearch is that although it has links to external databases such as HES, the anonymisation process in compiling the database means that there is no way to identify patients.

**THIN** is collaboration between two companies; In Practice Systems Ltd. (INPS), who developed Vision software used by General Practitioners in the UK to manage patient data, and Cegecim Healthcare Software.<sup>7</sup> THIN data collection started in 2003 and over 500 vision practices have so far joined the scheme. THIN data currently contain the electronic medical records of 11.1 million patients (3.7 million active patients). This covers 6.2% of the UK population.<sup>7</sup> In addition to the main consultations being recorded, the most patient data in THIN are linked to postcode-level area-based socioeconomic, ethnicity and environmental indices. The data are based on the patients' postcodes so that variables at ward level are available.<sup>8</sup> The patient is identified only by a code allocated by the GP system and cannot be identified outside the practice.

## AIM

The aim of this study was to conduct a bibliometric review to analyse the research outputs and the longitudinal growth in the number of publications that harness these three primary care databases; CPRD, QResearch and THIN from 2004 until 2013, and also to look at the growth of CPRD on its own from 1993 to 2013 as a case study.

## METHODS

To evaluate the impact of the three primary care databases (CPRD, QResearch and THIN), publications using data from each of the databases were extracted and analysed. For CPRD, we extracted publications from 1993 to 2013 using the Science Citation Index of the Thomson Scientific Institute for Scientific Information (Web of Science). Conference abstracts and posters were not included in the data. We used the same method to extract QResearch publications from 2004 to 2013. We obtained data for THIN from the Department of Primary Care and Population Health, University College London (UCL) and verified the data using the Web of Science. The data was provided in Excel format, which contained details of the author, title of article, the journal in which the article was published, article reference and the year of publication. Publications provided by UCL dated from 2004 to 2013.

The number of times publications were cited, their speciality area and the names of the journals they were published in were then extracted using the Web of Science for all three databases. The speciality areas of publications in all three databases were then categorised into four groups.

Speciality areas were classed as follows for all three databases.

- Speciality areas with  $\geq 50$  publications = Group 1.
- Speciality areas with  $< 50$  but  $\geq 20$  publications = Group 2.
- Speciality areas with publications  $< 20$  but  $\geq 10$  publications = Group 3.
- Speciality areas with  $<$  than 10 publications = Group 4.

All three databases were assessed for overall growth and their speciality areas were compared for similarities. To fairly compare the three databases, publication analysis had to focus on the overlapping period of data availability, which is 2004 to 2013. The publications in the three databases were compared, looking at the average number of publications per year, individual growth and cumulative growth over the 9-year period. The three 'most published in' journals were also recorded across the databases.

## RESULTS

### The CPRD database

1,140 publications categorised into 28 speciality areas were extracted for CPRD. Results represented in Table 1 show the number of CPRD publications from 1993 to 2013.

The largest speciality areas with  $> 50$  publications (Group 1) were made up from the following speciality areas: Pharmacology (284 publications), Health Services Research (71 publications) and Public Health (57 publications).

The average number of articles published per year over a 20-year period from 1993–2013 was 54, with the most listed in 2013 (144 publications). CPRD has shown consistent growth in publication output since its data started being used for research. As expected, there were relatively a few publications in the early years. Publications from 1993 to 1997 were 10, 9, 8, 7 and 17 publications, respectively. This rate picked

up in 1998 and 1999, with 30 articles in each year, and it has continued to grow since this point.

Between 2004 and 2013 (9-year period), the total number of publications listed in the CPRD database was 825, which shows significant growth in this period. 72.3% of the CPRD publications were published in the last 9 years. The highest number of publications was published in the *Pharmacoepidemiology and Drug Safety* journal, which represents 4.5% (52) of the publications in CPRD. Table 2 shows the journals that CPRD papers were most frequently published in.

### QResearch

Seventy-seven articles categorised into 13 speciality areas were published from studies conducted with QResearch data between 2004 and 2013 that were extracted from Web of Science. Results are listed in Table 3 below.

The average publications published per year over the 9-year period are around 8, with most articles published between 2011 (17 publications) and 2012 (20 publications).

The highest number of articles was published in the *British Journal of General Practice*, which represented 24.6% (19) of the publications in QResearch.

### THIN database

Three hundred-ninety four (394) articles categorised into 32 speciality areas from studies conducted with THIN data between 2004 and 2013 were extracted. Results are listed in Table 4.

The largest speciality areas with publications from THIN, with  $> 50$  publications (Group 1) were the following: Pharmacology (116) and General Medicine (51). The average number of articles published per year over the 9-year period was around 39, with the most publications listed in 2010 (87 publications). The *Pharmacoepidemiology and Drug Safety* journal published 21.3% (95) of publications. The remaining 78.7% of articles were published in journals that published between 1 and 14 articles. The majority of journals that published articles on THIN data only published once.

### Combined growth

Results represented in Table 5, and Figure 1 showed an increase in publications using data from all three databases. Over the 9-year period, publications for THIN and QResearch

**Table 1** Summary of search criteria used in the Web of Science

Database	Source of Original Data	Search Terms in Web of Science	Years Searched	Filtering Method
CPRD	Web of Science	Clinical Practice Research Database, CPRD and GPRD	1993–2013	Posters and reports were excluded
QResearch	Web of Science	QResearch	2004–2013	Posters and reports were excluded
THIN	UCL	Publication authors and year used as search criteria to verify	2004–2013	Posters and reports were excluded
HES	Web of Science	HES Secondary Care Database	1996–2014	Posters and reports were excluded

**Table 2 Annual number of publications published per database between 2004 and 2013**

Year	CPRD	THIN	QResearch	Total Publications
2004	70	2	2	74
2005	68	9	2	79
2006	61	14	2	77
2007	61	30	4	95
2008	72	45	7	124
2009	85	43	4	132
2010	69	87	9	165
2011	94	43	17	154
2012	101	72	20	193
2013	144	49	10	203
TOTAL	825	394	77	1296

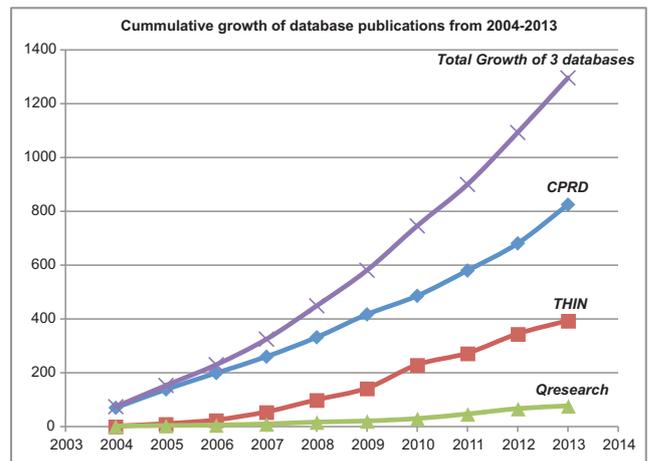
have slowly increased over time. CPRD publications have increased substantially in the last 4 years. All the three databases produced a combined total of 1,296 publications over the 9-year period (2004–2013) with CPRD representing 63.6% (825), THIN 30.4% (394) and QResearch 5.9% (77) of the total collective number of publications across these electronic health databases (EHDs). The growth across all three databases has been consistent over the 9-year period, with the highest number of publications listed in 2013.

The journal most published in, across the three databases, was the *Pharmacoepidemiology and Drug Safety* journal. As represented in Table 6, the most publications from CPRD and the THIN database occurred in this journal.

## DISCUSSION

This review looked at the publications using data from the three main UK primary care databases: CPRD, QResearch and THIN. Other databases derived from primary care data do exist; such as the Prescribing Analysis and Cost (database) and the Quality Management and Analysis System, but they have not been included in this review because they do not contain data based on individual patient records. A more recently developed database, ResearchOne, that derived from SystemOne was also excluded because it is only now being started to be used for research.

This review showed a large increase in studies conducted with data from all the three databases over time. The combined



**Figure 1 Cumulative growth of database publications from 2004–2013**

number of publications based on data from these three databases from 2004 to 2013 was 1,296 publications. The publications covered a large number of specialty areas, providing evidence of the widespread usage of this data. CPRD data was found to be used most commonly by researchers conducting studies in the Pharmacology specialty area, implying its importance for pharmaco-epidemiological and drug safety research and its management by the UK medicines regulator, the MHRA. This category accounted for 27% of the publications over the 20-year period reviewed for CPRD. QResearch attracted researchers who conducted more studies in the General and Internal Medicine specialty area. This accounted for 45% of the publications for QResearch. Finally, THIN database showed that researchers in the Pharmacoepidemiology and drug safety specialty area conducted more studies using its data, accounting for 22% of the publications generated from THIN data. Overall, publications from all the three databases showed that researchers from varying specialty areas showed a keen interest in the use of EHDs for research and, over time, this is bound to increase.

Some limitations do exist with this study because the publications obtained in each database varied in amount. The CPRD, being the largest of the three, contained more publications, over 90% more when compared to QResearch and 46.1% more than THIN. The THIN database when compared to QResearch contained 82% more publications. Some publications could also have been missed during extraction from the Web of Science. The details of THIN publications were primarily obtained from the Department of Primary Care and Population Health, UCL

**Table 3 Publications published in CPRD between 1993 and 2013**

CPRD Speciality Area Grouping	Number of Publications	Total Publications	% Publications
CPRD Group 1	≥50 publications	412	36.1
CPRD Group 2	<50 publications, ≥20	229	20.1
CPRD Group 3	<20 publications, ≥10	117	10.3
CPRD Group 4	<10 publications	382	33.5

**Table 4 Publications published from QResearch between 2004–2013**

THIN Speciality Area Grouping	Number of Publications	Total Publications	% Publications
General and Internal Group 1	≥50 publications	0	0
QResearch Group 2	<50 publications, ≥20	35	45.5
QResearch Group 3	<20 publications, ≥10	26	33.8
QResearch Group 4	<10 publications	16	20.8

and checked against the Web of Science. Also, the blanket categorisations (speciality areas) of the Web of Science were used, which may not always be reflective of a publication's true audience. Further, publications published in non-peer reviewed journals would not be accounted for in this work. Finally, another primary care database, ResearchOne, was not included in this study, as it has only recently been established.

In spite of these limitations, the growth in publications derived from these databases is clear. The use of EHDs in this manner remains invaluable and researchers are beginning to realise their benefits. The increase in publication for the combined databases from 74 in 2004 to 1296 in 2013 clearly indicates that this method of research is becoming more popular. It also highlights the importance of perfecting any limitations that this research method may present. Opportunities to use patient records for secondary uses are also on the increase with advances in technology allowing routinely collected data to be easily stored and shared.

Studies conducted in the past have also highlighted that EHDs such as CPRD will promote scientific productions in many ways.<sup>9</sup> For example, the clinical impact of these databases can be significant; namely, studies based largely on CPRD data and entirely from database research have contributed to the evidence for management, investigation and referral draft consultation document by the National Institute for Health and Care Excellence.<sup>3,10</sup> The wealth of information available in EHDs from primary care and their uses is invaluable and plays a key role in healthcare improvement in the UK. Studies that have provided useful information for many disease areas have used data available from both primary and secondary databases. The use of the data available in these databases has increased significantly over time and more recently, as seen in this review, has attracted more international interest. This is evident in the authorship of publications available from such studies. These studies have

provided researchers with great tools to help make the best decisions for the care of their patients.

Monitoring bodies have also been beneficiaries of the data contained in these databases because of the surveillance information that they provide. With the ability of CPRD to now link to other databases such as the HES, mortality statistics from ONS, cancer registry data from the National Cancer Intelligence Network, cardiovascular disease registry data from the Myocardial Ischaemia National Audit Project and Socioeconomic data at the Lower Layer Super Output Area level, the value of the database is expanding.<sup>11–14</sup>

It is also important to consider differences in the data quality between the primary care databases inherent to the computerised medical records (CMR) systems that contribute to them. Data contained within CPRD have historically been derived mainly from Vision/INPS, only recently taking some data from EMIS. The THIN database contains data exclusively from Vision/INPS, while QResearch uses only EMIS. Vision/INPS is a Problem Oriented Medical Record (POMR) software, meaning that it mandates the linkage of medical and drug information or visit consultations to a specific coded problem(s).<sup>15</sup> EMIS, however, is an Episode Oriented Medical Record (EOMR) software, meaning that this linkage is not obligated. It is widely accepted that POMR-based systems reduce inpatient coding variability, which assists in maintaining higher quality data.<sup>15</sup> This study has demonstrated that data from CPRD and THIN account for the vast majority of research outputs from primary care databases. Considering these primarily used POMR CMR data, it may be the case that researchers are aware of the potential data inaccuracies inherent to EOMR systems, such as EMIS, and thus preferentially elect to use data from CPRD and THIN databases.

The success of EHDs and the impact on research is completely dependent on the quality of the data that has been entered.<sup>16</sup> Accuracy, consistency and completeness of the data they contained have always been a challenge, one that

**Table 5 Percentage of publications published in THIN between 2004–2013**

THIN Speciality Area Grouping	Number of Publications	Total Publications	% Publications
THIN Group 1	≥50 publications	167	42.4
THIN Group 2	<50 publications but ≥ 20	75	19.0
THIN Group 3	<20 publications but ≥ 10	105	28.6
THIN Group 4	<10 publications	47	11.9

**Table 6 Top 3 journals most published in across all 3 databases from 2004–2013**

Journal	CPRD Publications	QResearch Publications	THIN Publications	Total Publications
Pharmacoepidemiology and Drug Safety journal	52	4	95	151
British Medical Journal	19	18	14	51
British Journal of General Practice	23	19	1	43

is still seeking a solution. Evidence for the accuracy and validity of the data from clinical databases is mixed and varies between clinical areas, the individual databases and the use of data.<sup>17–19</sup> A recent systematic review exemplified the low usage of EHDs to inform national healthcare guidelines.<sup>20</sup> Out of 25 guidelines included in the review, only 43 CPRD/GPRD studies were referenced, highlighting how electronic health records (EHRs) are a relatively untapped resource for informing evidence-based medicine (EBM).<sup>20</sup> This may in part be due to the widely accepted status quo that randomised controlled trials (RCTs) are the ‘gold standard’ data source for informing EBM. However, as exemplified in this study and the aforementioned review, research outputs from EHRs are rising, more rapidly so in recent years. It may be important to appreciate that in a primary care setting, ‘big data’ generated from EHR databases afford a number of advantages over RCT data, including better generalisability to real-life clinical settings and facilitation of increased autonomy over patients’ own healthcare data.

To increase the research outputs of EHR data, future research should concentrate on examining differences in the nature of data collection between databases particularly on a patient level, which will better inform how EHRs can be designed to answer clinical and population health questions. General Practitioners will need to be fully committed to reporting data, and increasing the quality of data reported and ensuring patient’s privacy are well protected by being gatekeepers of the data.<sup>21</sup>

## REFERENCES

- Gnani S and Majeed A. A user’s guide to data collected in primary care in England. Eastern Region Public Health Observatory, on behalf of Association of Public Health Observatories. 2006. Available at: <http://www.imperial.ac.uk/resources/579d8b09-c1c1-4026-a7be-c3e936ee9567>. Accessed 11 November 2014.
- The Clinical Practice Research Datalink. Available at: <http://www.cprd.com/home/>. Accessed 11 November 2014.
- Williams T, van Staa T, Puri S and Eaton S. Recent advances in the utility and use of the general practice research database as an example of a UK primary care data resource. *Therapeutic Advances in Drug Safety* 2012;3(2):89–99. Available at: <https://doi.org/10.1177/2042098611435911>. PMID:25083228; PMCID:PMC4110844.
- Kousoulis A. The CPRD and the RCGP: building on research success by enhancing benefits for patients and practices. *British Journal of General Practice* 2015;65(631):54–55.
- Williams T, van Staa T, Puri S and Eaton S. Recent advances in the utility and use of the general practice research database as an example of a UK primary care data resource. *Therapeutic Advances in Drug Safety* 2012;3(2):89–99. Available at: <https://doi.org/10.1177/2042098611435911>. PMID:25083228; PMCID:PMC4110844.
- What is QResearch? Available at: <http://www.qresearch.org/SitePages/What%20is%20QResearch.aspx>. Accessed 17 November 2014.
- THIN database. Available at: <http://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/database>. Accessed 1 December 2014.

## CONCLUSION

Based on the review of publications that harness data in CPRD, QResearch and THIN, there is strong evidence that these electronic health care databases are promoting scientific research. The growth in publications has shown that researchers are now conducting more studies using these databases and are beginning to realise their full potential. To continue to promote academic research, General Practitioners will need to continue to provide complete and accurate data; set standards will also need to be provided to General Practitioners to encourage enthusiasm and willingness to enter the required data; and public support encouraged for the continued use of these databases for research.

## Acknowledgement

Imperial College London is grateful for support from the North West London NIHR Collaboration for Leadership in Applied Health Research and Care. Fahmida Mannan and Zain Chaudhry contributed equally.

## Conflicts of Interest

Azeem Majeed has received funding for research on electronic patient records from the Medical Research Council, Engineering and Physical Sciences Research Council, NIHR, Wellcome Trust, Commonwealth Fund and Department of Health. Antonis Kousoulis is an employee of the CPRD.

8. Thin database description background. Available at: [http://www.itmat.upenn.edu/thin\\_bg.shtml](http://www.itmat.upenn.edu/thin_bg.shtml). Accessed 14 December 2014.
9. Chen YC, Wu JC, Haschler I, Majeed A, Chen TJ and Wetter T. Academic impact of a public electronic health database: bibliometric analysis of studies using the general practice research database. *PLoS One* 2011;6(6):e21404. doi: 10.1371/journal.pone.0021404.
10. Kousoulis AA, Rafi I and de Lusignan S. The CPRD and the RCGP: building on research success by enhancing benefits for patients and practices. *British Journal of General Practice* 2015;65(631):54–55.
11. National Institute for Health and Care Excellence. Suspected cancer: recognition and management of suspected cancer in children, young people and adults. 2015. Available at: <http://www.nice.org.uk/guidance/indevelopment/gid-cgwave0618> Accessed 14 December 2014.
12. Eaton S, Setakis E, Williams TW and van Staa TP. Linking primary care data (UK GPRD) to hospital records (HES). *Pharmacoepidemiology and Drug Safety* 2010;19:S195.
13. Setakis E, Gallagher A, Williams TW and van Staa TP. Linking UK death certificate data to UK primary care data (GPRD) to obtain cause of death. *Pharmacoepidemiology and Drug Safety* 2010;19:S83.
14. Boggon R, van Staa TP, Chapman M, Gallagher AM, Hammad TA and Richards MA. Cancer recording and mortality in the general practice research database and linked cancer registries. *Pharmacoepidemiology and Drug Safety* 2013;22(2):168–75. Available at: <https://doi.org/10.1002/pds.3374>. PMID:23239282.
15. de Lusignan S, Liaw ST, Dedman D, Khunti K, Sadek K and Jones S. An algorithm to improve diagnostic accuracy in diabetes in computerised problem orientated medical records (POMR) compared with an established algorithm developed in episode orientated records (EOMR). *Journal of Innovation in Health Informatics* 2015;22(2):255–64. doi: 10.14236/jhi.v22i2.79. Available at: <https://doi.org/10.14236/jhi.v22i2.79>.
16. Herrett E1, Thomas SL, Schoonen WM, Smeeth L and Hall AJ. Validation and validity of diagnoses in the general practice research database: a systematic review. *British Journal of Clinical Pharmacology* 2010;69(1):4–14.
17. Aylin P, Bottle A and Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ* 2007;334(7602):1044. Available at: <https://doi.org/10.1136/bmj.39168.496366.55>. PMID:17452389; PMCID:PMC1871739.
18. Cook JA and Collin GS. The rise of big clinical databases. *British Journal of Surgery* 2015;102(2):e93–e101.
19. Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ* 2013;346:f2350.
20. Oyinlola JO, Campbell J and Kousoulis AA. Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidances. *BMC Health Services Research* 2016;16:299. doi: 10.1186/s12913-016-1562-8. Available at: <https://doi.org/10.1186/s12913-016-1562-8>.
21. Gibson-White A and Majeed A. The Wellcome Trust Report: moving forward the use of general practice electronic patient records for research. *Informatics in Primary Care* 2009;17(3):141–42. Available at: <https://doi.org/10.14236/jhi.v17i3.727>.