

BMJ Open Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: a cross-sectional population study

Evangelos Kontopantelis,^{1,2} Richard John Stevens,³ Peter J Helms,⁴ Duncan Edwards,⁵ Tim Doran,⁶ Darren M Ashcroft^{1,7}

To cite: Kontopantelis E, Stevens RJ, Helms PJ, *et al*. Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: a cross-sectional population study. *BMJ Open* 2018;**8**:e020738. doi:10.1136/bmjopen-2017-020738

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-020738>).

Received 20 November 2017
Revised 30 November 2017
Accepted 4 December 2017



For numbered affiliations see end of article.

Correspondence to Professor Evangelos Kontopantelis; e.kontopantelis@manchester.ac.uk

ABSTRACT

Objectives UK primary care databases (PCDs) are used by researchers worldwide to inform clinical practice.

These databases have been primarily tied to single clinical computer systems, but little is known about the adoption of these systems by primary care practices or their geographical representativeness. We explore the spatial distribution of clinical computing systems and discuss the implications for the longevity and regional representativeness of these resources.

Design Cross-sectional study.

Setting English primary care clinical computer systems.

Participants 7526 general practices in August 2016.

Methods Spatial mapping of family practices in England in 2016 by clinical computer system at two geographical levels, the lower Clinical Commissioning Group (CCG, 209 units) and the higher National Health Service regions (14 units). Data for practices included numbers of doctors, nurses and patients, and area deprivation.

Results Of 7526 practices, Egton Medical Information Systems (EMIS) was used in 4199 (56%), SystemOne in 2552 (34%) and Vision in 636 (9%). Great regional variability was observed for all systems, with EMIS having a stronger presence in the West of England, London and the South; SystemOne in the East and some regions in the South; and Vision in London, the South, Greater Manchester and Birmingham.

Conclusions PCDs based on single clinical computer systems are geographically clustered in England. For example, Clinical Practice Research Datalink and The Health Improvement Network, the most popular primary care databases in terms of research outputs, are based on the Vision clinical computer system, used by <10% of practices and heavily concentrated in three major conurbations and the South. Researchers need to be aware of the analytical challenges posed by clustering, and barriers to accessing alternative PCDs need to be removed.

INTRODUCTION

Primary care in the UK has been almost fully computerised since the early 1990s.¹ Following the introduction of partial subsidies for the

Strengths and limitations of this study

- Cross-sectional analysis of all clinical computer systems used in English primary care, in August 2016.
- Data allowed a detailed description of regional use of each clinical computer system at the Clinical Commissioning Group level, and the discussion of implications for UK primary care databases (PCDs).
- Although regional presence of a clinical computer system cannot be equated to contribution to a PCD, since contributing practices are anonymised, inferences on the regional representativeness of UK PCDs are still possible.

purchase of clinical computer systems in 1998 and full subsidies in 2003 (in anticipation of the implementation of a national pay-for-performance programme) UK primary care became fully computerised.^{2–3} Interoperability requirements of the National Health Service (NHS) led to the universal adoption of a loosely hierarchical clinical coding system, known as Read codes,⁴ which is due to be replaced in April 2018 by a multi-hierarchical coding system (SNOMED CT).⁵ Various commercial providers were permitted to enter the market for clinical computer systems, resulting in numerous different systems with varying interfaces, mechanisms and implementations of Read code usage.⁶ By 2010–2011, seven clinical computer systems were consistently active in England, holding 99% of the market share: EMIS systems (LV and PCS) were active in 54.7% of practices, followed by Vision V.3 (18.1%) and SystemOne (17.8%), with the remaining 9.4% held by other systems (Synergy, Practice Manager, Premiere and the then newly launched EMIS Web).²

The uniformity and interoperability standards have facilitated the creation of large repositories of primary care electronic health records (EHRs), which contain the complete primary care records of patients attending general practices in the UK. The secondary use of these EHRs by researchers, both within and outside the UK, has been increasing exponentially,⁷ and they have provided insights in numerous research areas, including real-world effectiveness, adverse events, resource utilisation, condition prevalence and incidence, quality of care and policy interventions.^{8,9} Several EHR databases exist, maintained by the different clinical computer system providers, drawing data from practices using their systems that have agreed to make patient data available for secondary use. The four largest EHR databases (hereafter primary care databases or PCDs) in terms of numbers of patient records are the Clinical Practice Research Datalink (CPRD), The Health Improvement Network (THIN), QResearch and ResearchOne.

The CPRD (formally General Practice Research Database) was established in 1987 and has been owned by the Secretary of State for Health since 1994. In May 2017, the CPRD covered approximately 8% of the UK population, with 718 contributing general practices and over 17 million total patients (historical and current). The CPRD primarily collects data from Vision practices, although it is currently undergoing an expansion to include EMIS practices, and a future expansion to cover SystemOne practices is planned. THIN was established in 2003 as a collaboration between the company owning Vision (In Practice Systems) and the CSD Medical Research Group (now Quintiles IMS). In April 2015, THIN reported covering 6% of the UK population, with 562 practices and 11 million total patients. There is a considerable overlap (around 60%) between CPRD and THIN practices, which has implications for studies wishing to replicate findings between different databases.¹⁰ QResearch collects data from practices using EMIS systems and is the biggest PCD, with approximately 1500 practices in 2017, covering a population of more than 22 million patients.¹¹ ResearchOne is a collaboration between the provider of SystemOne, The Phoenix Partnership (TPP) and the University of Leeds, reporting 28 million (primary and secondary care) records and 423 practices in 2017.

The geographical coverage of PCDs is dependent on the location of practices using the parent clinical computer system, which is in turn dependent on historical patterns of market penetration by the software suppliers and system uptake by general practices. Geographical representativeness is an important prerequisite if analysts are to generalise PCD findings to the whole of England and the UK, which is what routinely happens in practice. This is due to great regional variability across England in terms of population characteristics (primarily: age, ethnicity and deprivation),¹² or even regional variation in hard outcomes. For example, a persistent mortality divide between North and the South of England has existed since the middle

of the previous century,¹³ while, more recently, much higher mortality rates were observed for young adults in the North of England.¹⁴ There is also regional variation in the organisation and productivity of health services in England,^{15,16} which could have important implications for the generalisability of health services research with the use of regionally unrepresentative PCDs. The aim of this paper is to describe the regional distribution of clinical computer systems in English primary care, evaluate the implications of the current picture of representativeness and provide some insight into the sustainability of existing PCDs.

METHODS

Data

Clinical computer system information was obtained from NHS Digital after direct communication, for August 2016. Primary care workforce and patient information as of 30 September 2016 was downloaded from the NHS Digital website.¹⁷ At the practice level, information was available on geography (Clinical Commissioning Group (CCG) and NHS region), patient list size by age groups, and numbers and full-time equivalent for general practitioners (GPs) and nurses. Deprivation was quantified using the 2015 release of the Index of Multiple Deprivation (IMD), a complete aggregate measure widely used to quantify area deprivation, attributed to the practice location.¹⁸ Spatial coordinates for NHS organisational units in 2016 were obtained from the Office for National Statistics open geography portal.¹⁹ We focused on two organisational levels, the lower CCGs with 209 units, and the higher NHS regions with 14 units.

Analyses

For all aspects of data manipulation and analysis we used Stata V.14.1. Whenever medians are reported, we also report the 25th and 75th centiles. Spatial maps were plotted using the *spmap* command.²⁰ Practice-level data were aggregated by clinical computer system, to provide information on all patients, patients aged ≥ 75 years, GPs and nurses, practice location deprivation and list size. Counts for each clinical computer system, by NHS region, were also calculated. Spatial graphs at the CCG level, with additional information on NHS regions, were plotted for the three most popular clinical computer systems, to provide a visual guide in regional distribution and representativeness.

RESULTS

System information was missing for 49 (0.7%) of 7526 general practices. EMIS systems were used in 4199 practices (56%), with all but 23 of these using EMIS Web. SystemOne was used in 2552 (34%), Vision in 636 (9%) and Evolution in 90 (1%) practices. Patterns of area deprivation, based on the locations of general practices, were similar across all systems. SystemOne practices tended to

Table 1 Regional distributions of systems and the characteristics of their respective general practices*†‡§

	EMIS¶	SystmOne	Vision V.3	Evolution
Aggregates (%)				
Number of practices	4199 (56)	2552 (34)	636 (9)	90 (1)
Number of patients	32 191 392 (56)	20 199 414 (35)	4 601 205 (8)	629 166 (1)
Number of GPs	18 675 (57)	11 160 (34)	2433 (7)	393 (1)
Medians (25th and 75th centiles)				
IMD 2015**	22.2 (12.1, 37.4)	22.5 (12.8, 36.8)	22.4 (12.3, 37.0)	22.7 (14.4, 31.0)
List size	6833 (4257, 10 094)	7080 (4214, 10 553)	6279 (3988, 9759)	6222 (4743, 9121)
Patients aged ≥75 years	476 (240, 823)	524 (256, 895)	455 (225, 710)	592 (400, 924)
Means (SD)				
All GPs	5.1 (3.4)	5.1 (3.7)	4.5 (3.1)	5.5 (2.9)
Female GPs	2.8 (2.4)	2.7 (2.5)	2.3 (2.1)	2.7 (2.1)
GPs aged <40 years	1.7 (1.9)	1.6 (2.0)	1.4 (1.7)	1.5 (1.6)
GPs aged 40–54 years	2.3 (1.9)	2.4 (2.1)	2.0 (1.8)	2.8 (1.9)
GPs aged ≥55 years	1.0 (1.0)	1.0 (1.0)	1.1 (1.0)	1.2 (1.1)
All nurses	3.1 (2.3)	3.6 (2.6)	2.8 (1.9)	3.4 (1.6)
Regional counts, NHS regions (%)				
Wessex	164 (55)	113 (38)	17 (6)	4 (1)
London	917 (68)	254 (19)	182 (13)	1 (0)
Yorkshire and the Humber	186 (25)	544 (74)	5 (1)	0 (0)
Cumbria and the North East	270 (59)	172 (38)	12 (3)	0 (0)
Cheshire and Merseyside	353 (92)	19 (5)	8 (2)	2 (1)
North Midlands	260 (54)	216 (45)	2 (0)	2 (0)
West Midlands	496 (76)	96 (15)	58 (9)	0 (0)
Central Midlands	156 (28)	378 (69)	16 (3)	0 (0)
East	112 (21)	413 (77)	4 (1)	4 (1)
South West	225 (59)	86 (22)	7 (2)	65 (17)
South East	303 (56)	96 (18)	145 (27)	1 (0)
South central	227 (55)	129 (31)	57 (14)	3 (1)
Greater Manchester	310 (65)	36 (8)	123 (26)	8 (2)
Lancashire	220 (100)	0 (0)	0 (0)	0 (0)

*Data for August 2016 (clinical system) and September 2016 (GMS data).

†System information not available for 49 (0.65%) of 7526 practices.

‡All GP numbers exclude locums.

§SystemOne provided by TPP, Vision (version 3) provided by In Practice Systems, Evolution provided by Microtest.

¶EMIS includes Web (4176 practices), LV (19 practices) and PCS (4 practices).

**Index of Multiple Deprivation (higher score implies higher levels of deprivation); details available in the 2015 technical report of the English Indices of Deprivation.¹⁸

GMS, General Medical Services; GPs, general practitioners; IMD, Index of Multiple Deprivation; NHS, National Health Service.

be larger (median of 7080 patients), followed by EMIS (6833), Vision (6279) and Evolution (6222).

Great regional variability in system usage was observed both at the NHS region level (table 1) and CCG level (figures 1–3). EMIS is present in all but 18 of the 209 CCGs (91.4%), with a much stronger presence in the West of England, London and the South. SystmOne is present in 120 CCGs (57.4%), and is mainly active in the East and some regions in the South. Vision, although with a much lower market share than SystmOne, is still used in

96 CCGs (45.9%), mainly in London, the South, Greater Manchester and Birmingham. Evolution is only present in 18 CCGs (8.6%) and is primarily used in the South West.

DISCUSSION

High regional variability exists in the use of different clinical computer systems in English primary care, which should be a consideration when utilising primary care

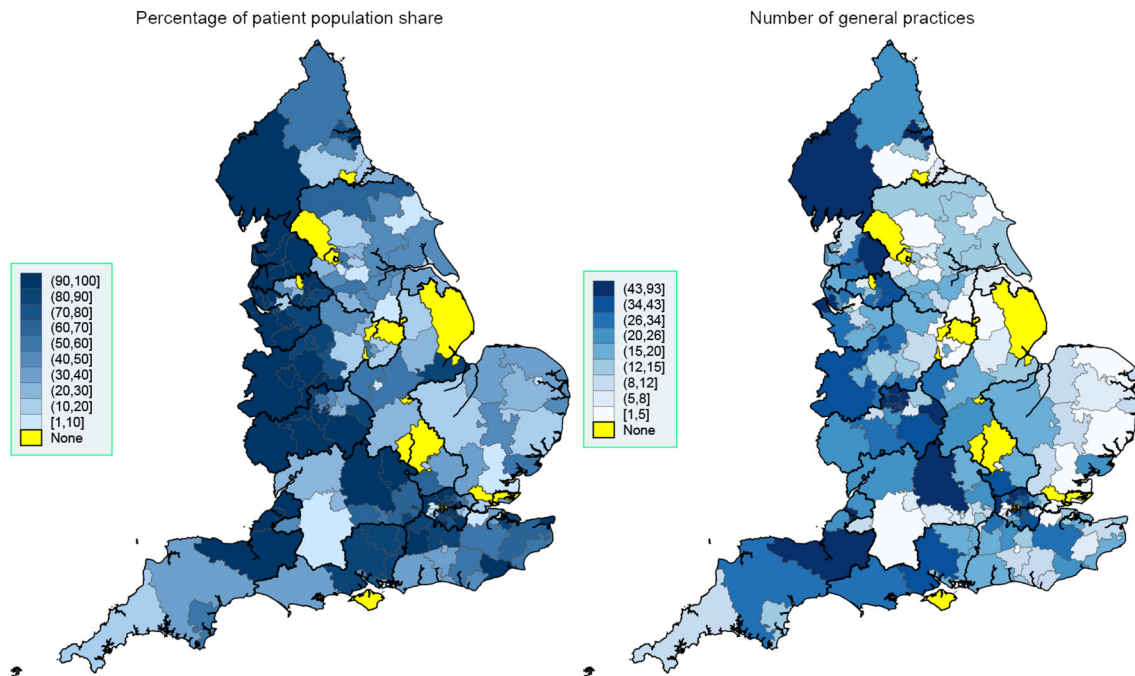


Figure 1 Spatial map at the CCG level, September 2016: EMIS. Thicker border lines correspond to the 14 NHS regions, left graph uses equidistant class breaks; right graph uses class breaks based on distribution of variable of interest, with each class having approximately the same number of spatial polygons (CCGs). CCG, Clinical Commissioning Group; NHS, National Health Service.

electronic health databases based on this population in the future, especially if effect heterogeneity (or other forms of heterogeneity) is context relevant. For example, drawing nationwide conclusions in health services organisation would be more problematic than identifying

medication side effects. EMIS Web is by far the most widely used clinical computer system and therefore QResearch is the most nationally representative single database—potentially able to collect data from almost all English CCGs. SystemOne has a very strong presence in many parts

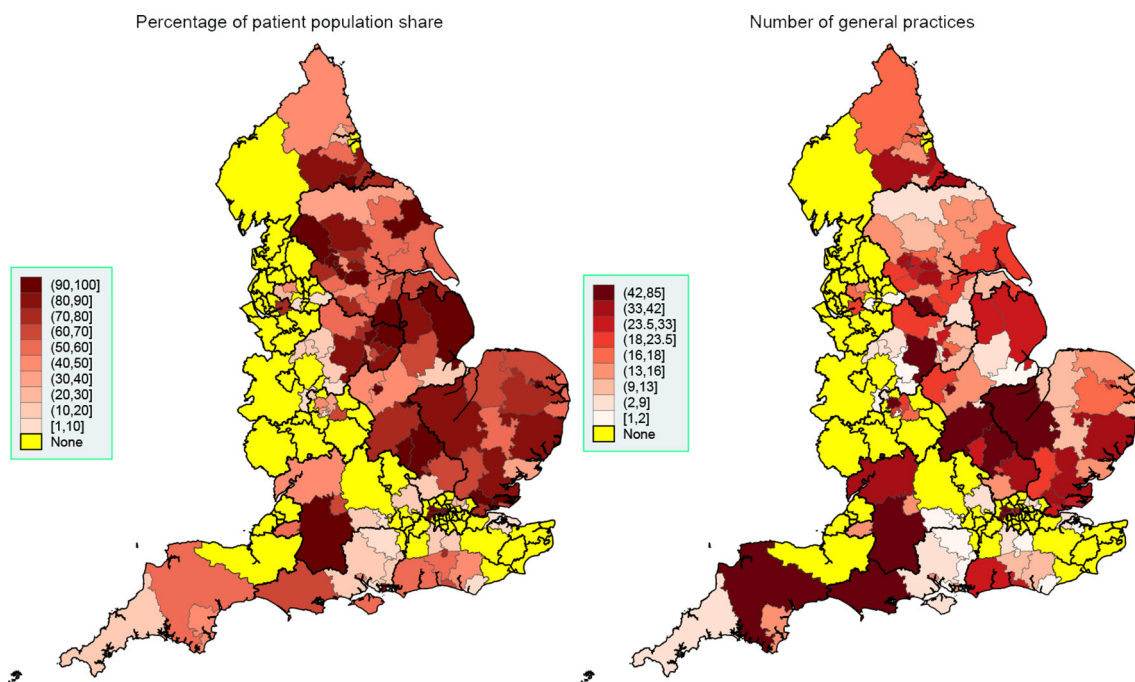


Figure 2 Spatial map at the CCG level, September 2016: SystemOne. Thicker border lines correspond to the 14 NHS regions, left graph uses equidistant class breaks; right graph uses class breaks based on distribution of variable of interest, with each class having approximately the same number of spatial polygons (CCGs). CCG, Clinical Commissioning Group; NHS, National Health Service.

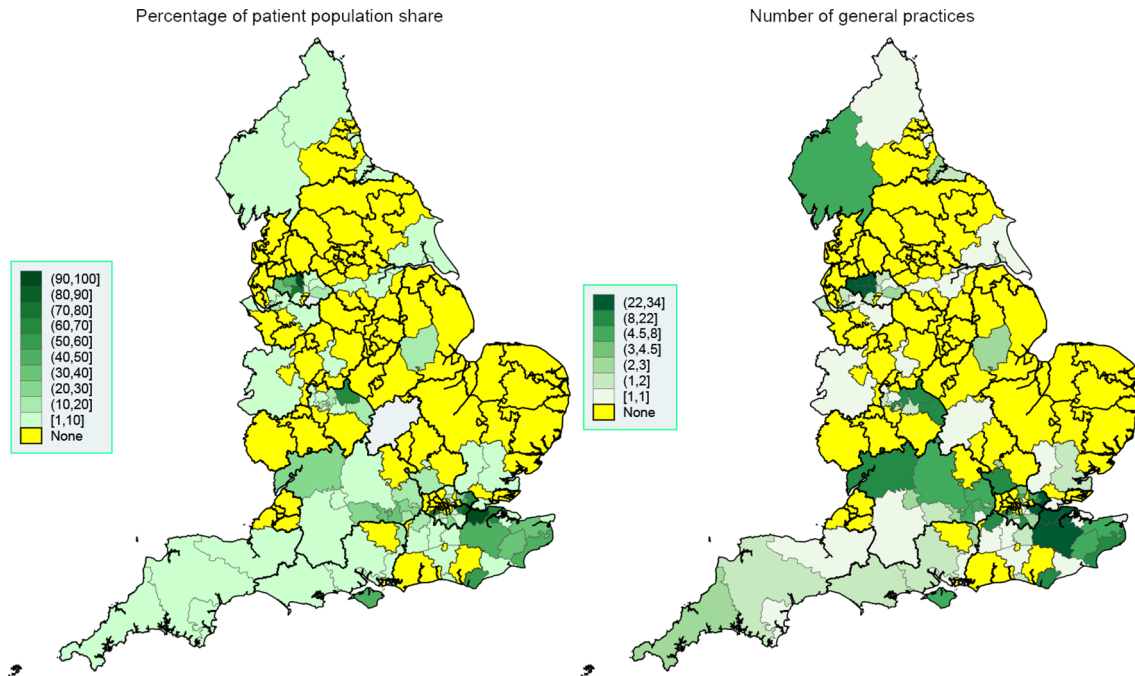


Figure 3 Spatial map at the CCG level, September 2016: Vision. Thicker border lines correspond to the 14 NHS regions, left graph uses equidistant class breaks; right graph uses class breaks based on distribution of variable of interest, with each class having approximately the same number of spatial polygons (CCGs). CCG, Clinical Commissioning Group; NHS, National Health Service.

of England, but no presence in many CCGs in the North West, West Midlands, London and South East. The ResearchOne database is therefore unable to capture data from many regions. Finally, Vision is the most geographically restricted of the three major clinical computer systems, with relatively few practices heavily concentrated in three conurbations and the South. The CPRD and THIN databases are therefore currently unable to provide comprehensive coverage of large parts of the country, particularly in the North and East of England.

Strengths and limitations of the study

The main strength of this study is the use of numerous national administrative datasets of high data quality, allowing us to obtain a complete picture for the whole of England. The main weakness of the work is the fact we cannot equate the regional presence of a clinical system to active contribution to a primary care database—not all practices contribute data and contributing practices are anonymised—and we have therefore discussed potential contribution instead. Additional information on currently registered patients would have been relevant, but is not routinely available for non-users of the resources (but can be deduced by users).

Findings and implications

The current picture of clinical system usage in English primary care is very different to what was reported for 2011.² Although EMIS is still the biggest provider and has retained its market share (56% in both 2011 and 2016), its LV and PCS systems which dominated the market in 2011 are hardly used anymore, with almost all practices

having transitioned to the Web system. The use of TPP's SystemOne has increased from 18% to 34%, while that of Vision by In Practice System has halved (from 18% to 9%). Many providers that were present in 2011 have subsequently withdrawn from primary care, with the exception of Microtest's Evolution (transitioned from Practice Manager). If the current trend continues, English primary care will be completely dominated by EMIS Web and SystemOne in the next 5–10 years, and access to both of these systems would ensure almost complete coverage for England.

The trend for primary care convergence to two clinical systems has implications for the future of PCDs and the research findings based on them. CPRD and THIN will need to adapt very quickly and include EMIS and/or SystemOne practices in their processes. Given that the CPRD and THIN are the two most widely used primary care databases in clinical research, losing them altogether, as happened with the DIN-LINK database,²¹ would be a severe setback for the research community. As of 20 July 2017, a PubMed search identified 1782 published papers linked to the CPRD (886 in the last 5 years), 471 linked to THIN (303 in the last 5 years), 71 linked to QResearch (32 in the last 5 years) and two to ResearchOne (both in the last 5 years). Although not exhaustive, this search indicates the large variability across databases in terms of scientific contribution, demonstrating that the most accessible and productive databases are the ones at immediate risk.

Within the CPRD, there are clear actions towards future-proofing the resource, in light of the deterioration of the Vision market share. A large number of EMIS

practices are already contributing data to the resource, but differences in the data format (compared with the standard Vision format) has prevented their immediate release along with Vision data, while it was not possible to link the EMIS data to other data sets. Nevertheless, a major transformation in processes is being undertaken which will allow the release of both Vision and EMIS data as standard, within 2018. In addition, the recruitment of EMIS practices continues, with over 150 practices having joined the CPRD in the last 12 months.

Users of the UK PCDs need to be aware of the generalisability issues we described, and consider if there are any risks relevant to their studies. Generalisability (external validity) should be discussed as standard in such work and is listed as an item (no. 21) in both the Strengthening the Reporting of Observational Studies in Epidemiology and Reporting of studies Conducted using Observational Routinely collected health Data statements.^{22 23} The context is important here, and regional representativeness may be less relevant for clinical questions but more relevant for health services research. Sensitivity analyses on a more representative group of practices, obtained through deterministic sampling and existing software,²⁴ can also be used to strengthen findings.²⁵ However, the strong clustering of clinical systems within CCGs, largely driven top-down from CCGs to general practices, limits the usefulness of such sampling approaches.

CONCLUSIONS

The geographical representativeness of primary care databases varies enormously, and the two most used databases in the UK, the CPRD and THIN, were in 2016 the least representative of the major databases due to the quickly diminishing market share of the clinical computer system providing their data (Vision). The existence of these databases is under threat, and urgent action is required to allow data collection from at least one of the two dominant clinical systems (EMIS Web and SystemOne). CPRD has recognised this, and has recently negotiated access to data held by EMIS practices, and is due to operationalise this data by 2018. In addition, development and access barriers that have restricted publication outputs from data drawn from EMIS (QResearch) and SystemOne (ResearchOne) practices urgently need to be overcome if the confidential use of NHS patient data is to continue driving research that directly informs patient safety, management and health services policy.

Author affiliations

¹NHRC School for Primary Care Research, University of Manchester, Manchester, UK

²Division of Population Health, Health Services Research and Primary Care, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

³Nuffield Department of Health Care Sciences, University of Oxford, Oxford, UK

⁴The Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

⁵Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, UK

⁶Department of Health Sciences, University of York, York, UK

⁷Division of Pharmacy and Optometry, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

Acknowledgements We thank the Office of National Statistics and NHS Digital for the wealth of information they have collected and systematically organised, which made this study possible.

Contributors EK designed the study, extracted the data from all sources, performed the analyses and drafted the first version of the manuscript. RJS, PJH, DE, TD and DMA critically edited the manuscript. EK is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Funding MRC Health eResearch Centre Grant MR/K006665/1 supported the time and facilities of EK.

Disclaimer The views expressed are those of the authors and not necessarily those of the MRC.

Competing interests EK, RJS, PJH and DE are members of the Independent Scientific Advisory Committee (ISAC) for MHRA database research:

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The data used in this study are freely available and the authors are happy to share an organised and cleaned final dataset.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Millman A, Lee N, Brooke A. ABC of medical computing. computers in general practice--I. *BMJ* 1995;311:800-2.
2. Kontopantelis E, Buchan I, Reeves D, et al. Relationship between quality of care and choice of clinical computing system: retrospective analysis of family practice performance under the UK's quality and outcomes framework. *BMJ Open* 2013;3:e003190.
3. Roland M. Linking physicians' pay to the quality of care--a major experiment in the United Kingdom. *N Engl J Med* 2004;351:1448-54.
4. de Lusignan S, Chan T. The development of primary care information technology in the United Kingdom. *J Ambul Care Manage* 2008;31:201-10.
5. NHS England. The Terminology roadmap for the NHS: recommendations for the standardisation committee for care information. 2014:1-35 http://www.england.nhs.uk/iscg/wp-content/uploads/sites/4/2014/09/20140827_SCCI_Item_9_Report_v3.pdf.
6. Tai TW, Anandarajah S, Dhoul N, et al. Variation in clinical coding lists in UK general practice: a barrier to consistent data entry? *Inform Prim Care* 2007;15:143-50.
7. Springate DA, Kontopantelis E, Ashcroft DM, et al. ClinicalCodes: an online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. *PLoS One* 2014;9:e99825.
8. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ* 2003;326:1070.
9. Vezyridis P, Timmons S. Evolution of primary care databases in UK: a scientometric analysis of research output. *BMJ Open* 2016;6:e012785.
10. Carbonari DM, Saine ME, Newcomb CW, et al. Use of demographic and pharmacy data to identify patients included within both the Clinical Practice Research Datalink (CPRD) and The Health Improvement Network (THIN). *Pharmacoepidemiol Drug Saf* 2015;24:999-1003.
11. Hippisley-Cox J, Coupland C. Development and validation of risk prediction equations to estimate survival in patients with colorectal cancer: cohort study. *BMJ* 2017;357:j2497.
12. Doran T, Drever F, Whitehead M. Is there a north-south divide in social class inequalities in health in Great Britain? Cross sectional study using data from the 2001 census. *BMJ* 2004;328:1043-5.
13. Hacking JM, Muller S, Buchan IE. Trends in mortality from 1965 to 2008 across the English north-south divide: comparative observational study. *BMJ* 2011;342:d508.

14. Buchan IE, Kontopantelis E, Sperrin M, *et al.* North-south disparities in english mortality 1965-2015: longitudinal population study. *J Epidemiol Community Health* 2017;71:928–36.
15. Appleby J, Raleigh V, Frosini F, *et al.* *Variations in health care: the good, the bad and the inexplicable*: King's Fund, 2011.
16. Bojke C, Castelli A, Street A, *et al.* Regional variation in the productivity of the English national health service. *Health Econ* 2013;22:194–211.
17. NHS Digital. General and Personal Medical Services, England 2006-2016, as at 30 September, Experimental statistics. <http://content.digital.nhs.uk/article/2021/Website-Search?productid=24768&q=+General+and+Personal+Medical+Services&sort=Relevance&size=10&page=1&area=both#top>
18. Communities and Local Government. The english indices of deprivation 2015: technical report. government DfCaL. 2015 <https://www.gov.uk/government/publications/english-indices-of-deprivation-2015-technical-report>
19. Office for National Statistics. Open geography portal. 2017 <http://geoportal.statistics.gov.uk/>
20. IDEAS. SPMAP: Stata module to visualize spatial data [program]. 2017.
21. Carey IM, Cook DG, De Wilde S, *et al.* Developing a large electronic primary care database (Doctors' Independent Network) for research. *Int J Med Inform* 2004;73:443–53.
22. von Elm E, Altman DG, Egger M, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335:806–8.
23. Benchimol EI, Smeeth L, Guttman A, *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12:e1001885.
24. Kontopantelis E. A greedy algorithm for representative sampling: repsample in Stata. *J Stat Softw* 2013;55:1–19.
25. Kontopantelis E, Springate D, Reeves D, *et al.* Withdrawing performance indicators: retrospective analysis of general practice performance under UK Quality and outcomes framework. *BMJ* 2014;348:g330.