# BMJ Open

# Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile

Ana Correa,[1,2] William Hinton,[1] Andrew McGovern,[1] Jeremy van Vlymen,[1] Ivelina Yonova,[1,2] Simon Jones,[3] Simon de Lusignan[1,2]

[1]Department of Clinical and Experimental Medicine, University of Surrey, Guildford, UK
[2]Royal College of General Practitioners, London, UK
[3]Division of Healthcare Delivery Science, New York University, New York, NY, USA

**Correspondence to**
Professor Simon de Lusignan; s.lusignan@surrey.ac.uk

## ABSTRACT

**Purpose:** The Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) is one of the longest established primary care sentinel networks. In 2015, it established a new data and analysis hub at the University of Surrey. This paper evaluates the representativeness of the RCGP RSC network against the English population.

**Participants and method:** The cohort includes 1 042 063 patients registered in 107 participating general practitioner (GP) practices. We compared the RCGP RSC data with English national data in the following areas: demographics; geographical distribution; chronic disease prevalence, management and completeness of data recording; and prescribing and vaccine uptake. We also assessed practices within the network participating in a national swabbing programme.

**Findings to date:** We found a small over-representation of people in the 25–44 age band, under-representation of white ethnicity, and of less deprived people. Geographical focus is in London, with less practices in the southwest and east of England. We found differences in the prevalence of diabetes (national: 6.4%, RCPG RSC: 5.8%), learning disabilities (national: 0.44%, RCPG RSC: 0.40%), obesity (national: 9.2%, RCPG RSC: 8.0%), pulmonary disease (national: 1.8%, RCPG RSC: 1.6%), and cardiovascular diseases (national: 1.1%, RCPG RSC: 1.2%). Data completeness in risk factors for diabetic population is high (77–99%). We found differences in prescribing rates and costs for infections (national: 5.58%, RCPG RSC: 7.12%), and for nutrition and blood conditions (national: 6.26%, RCPG RSC: 4.50%). Differences in vaccine uptake were seen in patients aged 2 years (national: 38.5%, RCPG RSC: 32.8%). Owing to large numbers, most differences were significant ( p<0.00015).

**Future plans:** The RCGP RSC is a representative network, having only small differences with the national population, which have now been quantified and can be assessed for clinical relevance for specific studies. This network is a rich source for research into routine practice.

## Strengths and limitations of this study

- The Royal College of General Practitioners Research and Surveillance Centre network is representative of the national population on a variety of domains, both demographic and clinical, which support the network's suitability for real-world evidence research.
- Statistically significant differences have been found, due to the large numbers of the sample, but this profile has now quantified them, so that they can be assessed on a study-by-study basis for their clinical relevance.
- Some of the analysis compared data extracted using different methodology, which may have skewed the results.
- The network is comprised of volunteer practices, with better management of clinical conditions, which may introduce a bias to research into clinical outcomes.

## INTRODUCTION

Data from primary care sentinel networks is widely used in research and surveillance;[1] the Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) is one of the longest established.[2] These networks are of particular importance for conditions largely managed in primary care.[3] However, outputs may be biased if the sample is not representative of the monitored population, or if there is uncertainty about the denominator.[4] Estimating rates of disease is easiest from countries that have registration-based primary care systems,[5] though useful data can be obtained from health systems that do not require registration.[6]

Demographics and geographical spread are important, and many sentinel networks are involved in the surveillance of infectious conditions, such as influenza, and assessment of vaccine effectiveness. Different age-groups and differing levels of deprivation have different

rates of many illnesses, including rates of swab-confirmed influenza.[7] There are also disparities associated with ethnicity, though these are not consistent.[8–10]

For infectious diseases like influenza, geographic spread of sentinel practices is important, as well as consistent case ascertainment from year to year.[11] It is also important that the characteristics of the participating practices are representative.[12] In particular, when assessing vaccine effectiveness, vaccine uptake needs to be properly recorded within the database, in a way that is representative of the national population rates.[13]

Disease, prescription and vaccination patterns, as well as the underlying population demographics, are interrelated. Any assessment of quality requires an assessment of the rate of case ascertainment followed by an evaluation of the management of the condition.[14] Some sentinel networks have focused on the management of long-term conditions, including diabetes.[15] Chronic conditions, such as diabetes, benefit from improved recording of comorbidities and risk factors.

The RCGP RSC is a network of general practices, which extracts data from the computerised medical record systems of over 100 practices in England. The network established a weekly returns service in 1964, which has enabled prompt surveillance of infectious diseases and identification of epidemics; with influenza surveillance as a key priority for the network.

The characteristics of the RCGP RSC practice network have previously been compared with population-level data to ascertain the representativeness of the sample.[16] However, since the most recent report in 2009, there have been substantial changes within the network, including the commissioning in 2015 of an entirely new data and analytics hub at the University of Surrey.

We carried out this study to describe the representativeness of the RCGP RSC practice network comparing population demographics; geographical distribution; the prevalence and management of chronic conditions, and data completeness taking type 2 diabetes mellitus (T2DM) as an exemplar condition; medication prescribing; and vaccine uptake.

## METHODS

We compared four key areas of representativeness with the national population:
▶ Demographics: age, gender, ethnicity and deprivation;
▶ Geographical distribution;
▶ Prevalence and management of chronic conditions, and the data completeness of risk factors for T2DM patients;
▶ Medications prescribing rates and costs, and influenza vaccine uptake.

We also compared the demographic characteristics of the practices within the RCGP RSC network which provide nasopharyngeal virology swabbing specimens for Public Health England's (PHE) viral infection surveillance programme, with those that do not. The data for

this study was extracted from practices providing data uploads in March 2015 into the new RCGP RSC data upload system. RCGP RSC member practices are all volunteers who sign a written agreement to participate in surveillance and research. The practice upload includes all registered patients' coded data.

The data extracted by the RCGP RSC sentinel system is the principal primary care public health surveillance data used by PHE. All data extracted are pseudonymised as close to source as possible, and kept in a secure network database. Patients with flags in their records suggesting they have opted out of data sharing do not have their data retained or analysed. Data used for surveillance is part of direct patient care. However, all planned research use of this data requires review by research ethics committees prior to use.

### Population demographics

The RCGP RSC population was divided by gender, aggregated into the age bands used in the 2011 Census for England and Wales, and compared with national census data available from the Office for National Statistics. Age was calculated from the date of birth at the point of extraction. Similarly, patients were classified into five ethnic groups (according to census categories), and compared against national data. We used an updated version of an established ethnicity classification mapping to maximise the use of any ethnicity data.[17]

In the process of data extraction, patient postcodes were converted into Lower Super Output Areas (LSOAs),[18] and these were used to assign an Index of Multiple Deprivation (IMD) score to each patient.[19] Postcode data was then removed from the final extract to maintain pseudonymisation. We also compared the demographics (age, gender, ethnicity and IMD scores) of practices providing viral throat swabs for PHE with practices which do not.

### Geographic distribution

We used LSOA data, coordinates and map files provided by the Office for National Statistics geographical survey,[20] to display the geographical distribution of the practices. Using each GP practice postcode, we plotted this against a map of English Clinical Commissioning Group (CCG) regions, using the statistical software R, and the package maptools.[21]

### Chronic conditions

We compared the prevalence and quality of management of key chronic diseases. We used the data collected for a national pay-for-performance scheme, the Quality Outcomes Framework (QOF), to compare the prevalence of chronic conditions that are part of the scheme,[22] for RCGP RSC practices and at the national level. QOF quality points were used as a surrogate of the quality of care and management of these conditions.

Additionally, we selected diabetes as an exemplar condition to demonstrate the level of data quality found in

the database for chronic conditions. We assessed the level of completeness in data recording for a number of demographics (age, gender, deprivation) and risk factors (smoking status; systolic blood pressure; body mass index (BMI, or height and weight); ethnicity; cholesterol and glycaemic control), within the RCGP RSC database.

We only analysed coded data to define data completeness; while free-text data can be extracted, there are risks of extracting patient-identifiable data, so this data is not used within the RCGP RSC network. Coded data in primary care has been widely used in health outcomes research,[23] and has known strengths and limitations.[5]

### Medication prescribing and vaccine uptake
Prescription data for England was collated via the Prescription Cost Analysis 2014 report,[24] and these were compared with the RCGP RSC data for the same year. We calculated proportions for the most commonly prescribed British National Formulary (BNF) chapters (gastrointestinal, cardiovascular, respiratory, central nervous, endocrine, nutrition/blood and infections), and compared these with the national proportions.

The average cost per item of each BNF chapter was multiplied by the quantities prescribed to obtain costs per chapter, for both the RCGP RSC practices and national data. Additionally, we compared influenza vaccine uptake in the RCGP RSC database with national rates published by PHE,[25] focusing on four at-risk age groups (children aged 2, 3, 4 and over 65 years old).

### Statistical analyses
We used standard descriptive statistics to compare the RCGP RSC with national data. The $\chi^2$ test was used to compare proportions across most variables. Differences in the cumulative frequency distribution of IMD scores were assessed using the Kolmogorov-Smirnov test. The statistical software R was used for all tests of significance, and the p value was reported. Owing to the large number of statistical tests (69 tests), we applied a Bonferroni-Šidák correction[26 27] to a level of significance of 0.01, resulting in a new level of significance of 0.00015 (Equation 1). This version of the correction was used since all the tests are independent.
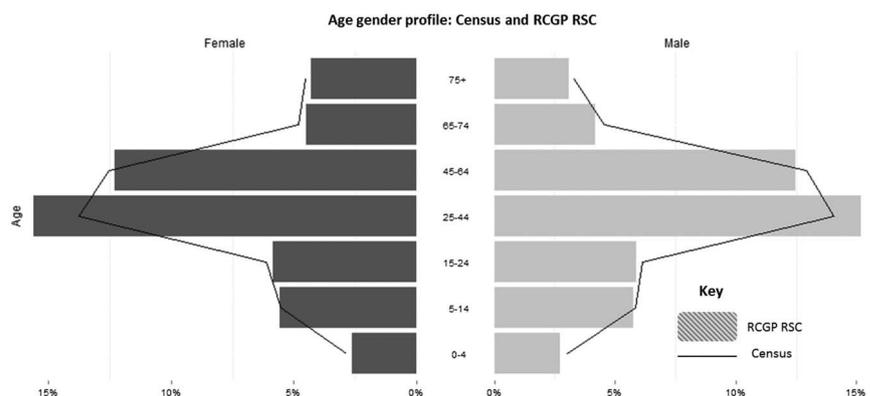
Bonferroni-Šidák equation

$$\alpha' = 1 - (1 - \alpha)^{1/m} \tag{1}$$

a′=new significance level, α=current significance level, m=number of tests.

## COHORT DESCRIPTION AND FINDINGS TO DATE
The cohort includes data from 1 042 063 patients registered at the time of data upload in one of the 107 participating practices.

### Population demographics
Data for age and gender was recorded for the complete RCGP RSC population. The age and gender distribution of the RCGP RSC population was similar to the census distribution (figure 1 and online supplementary table S1). There was a significantly higher proportion of both males and females in the 25–44 years age band when compared with the census, and a lower proportion of people in the 0–4 years age band. All differences were statistically significant; with the exception of people aged 65–74 years, men aged 45–64 years and women aged 5–14 years.

The ethnicity of 630 754 (61%) patients from the RCGP RSC cohort was identified from their medical records. The proportions of the five census ethnicity groups were similar to those from census data (see online supplementary figure S1 and table S2). The majority of patients in the RCGP RSC cohort were of white ethnicity (84.4%), similar to the census population (85.4%). Owing to the large numbers, the differences between the census and RCGP RSC data were significant, even when applying a Bonferroni-Šidák correction.

IMD scores were calculated for all patients included in the analysis, using LSOAs (see online supplementary figure S2). The mean IMD score for the RCGP RSC population was 19.8 (SD 0.00682), which was less deprived than the English population (mean IMD score 21.8; SD 0.00050). This is mostly due to a small over-representation of patients in the two least deprived deciles (see online supplementary figure S3). The Kolmogorov-Smirnov test showed that this difference was significant.

**Figure 1** Age-gender profile for the Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) compared with the census population.

**Figure 2** Geographical distribution of the Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) practices in England.



## Geographical distribution of the RCGP RSC practices
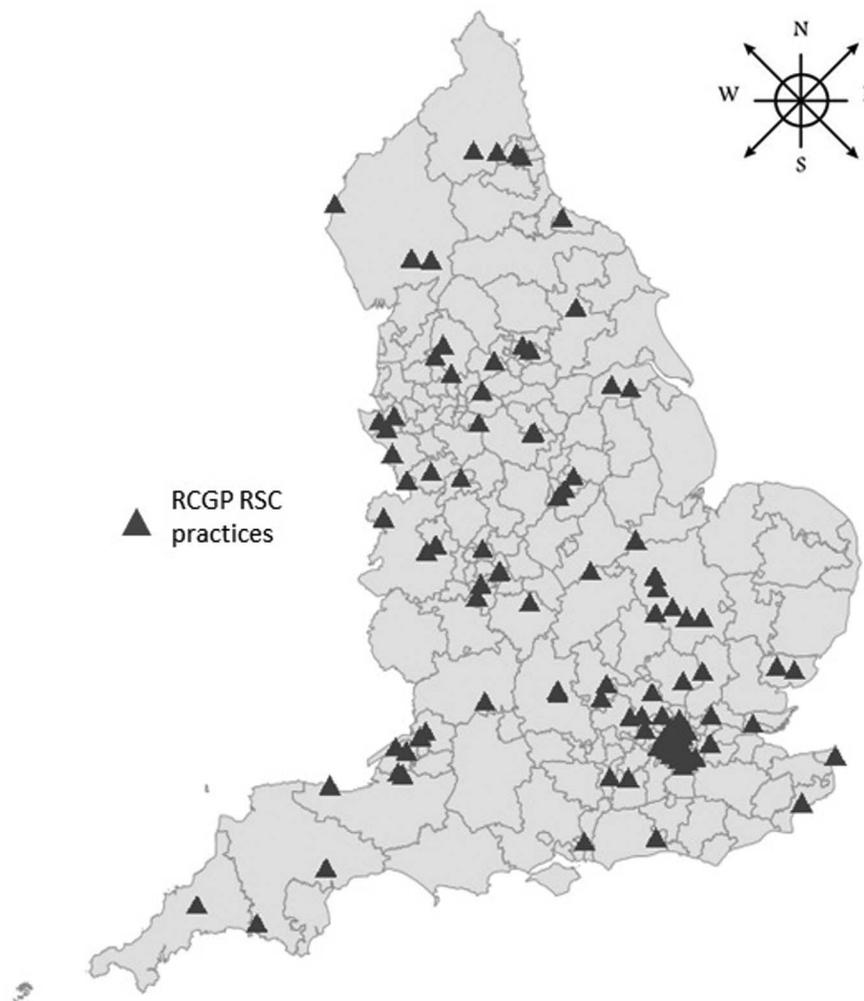
The demographic characteristics of the practices participating in the PHE swabbing programme were compared with non-participating practices. There was a significantly higher proportion of people in the 25–44 years age band for participating practices, and a lower proportion of people in the 65+ years band (see online supplementary figure S4). All differences were statistically significant, with the exception of men aged 15–24 years and women aged 5–14 years.

Participating practices had significantly less people in the white census group (75%) compared to non-participating practices (83%), and more people in ethnic minority groups (see online supplementary figure S5). Finally, the population in participating practices was more deprived (mean IMD score 21.5, SD 0.018) than that of the non-participating practices (mean IMD score 19.0, SD 0.010, see online supplementary figure S6), and the Kolmogorov-Smirnov test showed this to be significant.

### Geographic distribution

The RCGP RSC practices are broadly distributed across England, with a higher concentration of practices in London, and a slightly lower number in the southwest and east of England (figure 2). The distribution shows a number of practice clusters, probably due to effective recruitment strategies using the support of CCGs.

### Chronic conditions

The prevalence of common chronic diseases was similar to that reported nationally in the QOF scheme for chronic disease management (figure 3 and online supplementary table S3[i]). Diabetes (national: 6.4%, RCPG RSC: 5.8%), learning disabilities (national: 0.44%, RCPG RSC: 0.40%), obesity (national: 9.2%, RCPG RSC: 8.0%), and pulmonary disease (national: 1.8%, RCPG RSC: 1.6%) are slightly under-represented, while cardiovascular diseases (national: 1.1%, RCPG RSC: 1.2%) are slightly over-represented.

All differences were significant, with the exception of asthma, chronic kidney disease, depression, palliative care and stroke. The median percentage of QOF targets achieved by the RCGP RSC network (98.76%; IQR 2.64%) was higher than the national median percentage

---

[i]The denominators vary by condition, depending on the QOF definition (all population, population over 18 years, etc).
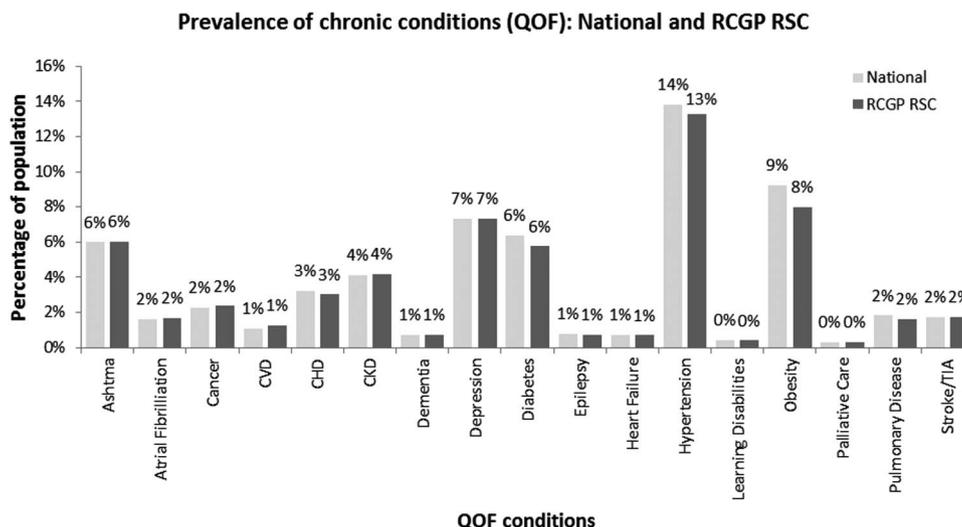
**Figure 3** Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC)-recorded prevalence of chronic conditions compared with all-English practices, using Quality Outcomes Framework (QOF) data.

(97.04%; IQR 5.45%), due to an over-representation of the network in the higher score deciles (see online supplementary figure S7).

Diabetes was taken as an exemplar condition to demonstrate the completeness of recorded data within the RCGP RSC database (table 1). This analysis was performed on patients who have complete data for the past 10 years. The recording of data related to cardiovascular risk factors ranges from 77% to 99% for the diabetic population, and from 26% to 69% for the non-diabetic population. Owing to the introduction of QOF targets, the data completeness related to risk factors or clinical

management is expected to be higher for patients with a chronic condition.[28]

The comparison with non-diabetic patients is provided to demonstrate the improvement in recording expected for patients with a chronic condition, compared to those without. Since the QOF targets incentivise accurate recording of chronic conditions diagnoses, and of continuous assessments of risk factors for patients with these conditions, primary care data is particularly well suited for longitudinal research into chronic conditions.

### Medication prescribing and vaccine uptake

Prescription items and costs by BNF chapters in the RCGP RSC dataset were compared with national data. The RCGP RSC had higher prescribing rates and costs for drugs within the infections chapter (national: 5.58%, RCPG RSC: 7.12%), while the practices in the network prescribed less drugs in the nutrition and blood chapter (national: 6.26%, RCPG RSC: 4.50%). This difference was significant (figure 4), and it was also reflected in the prescription costs (see online supplementary figure S8).

Influenza vaccine exposure rates within the RCGP RSC network were compared with the national rates published by PHE. Influenza vaccine uptake in the RCGP RSC network is similar to the national rates, although slightly smaller; these are shown in figure 5. The largest difference was seen in the group of patients aged 2 years (national: 38.5%, RCPG RSC: 32.8%), while the smallest differences were in the over 65 years olds (national: 73%, RCPG RSC: 71%).

### STRENGTHS AND LIMITATIONS
### Strengths
This study shows the representativeness of the RCGP RSC against nationally published data on a variety of domains, both demographic and clinical. Beyond the

**Table 1** RCGP RSC data completeness on key variables for patients with and without diabetes

| Variables | Diabetes | | Non-diabetes | |
|---|---|---|---|---|
| | Count | Per cent | Count | Per cent |
| Age | 39 559 | 100 | 977 445 | 100 |
| Gender | 39 559 | 100 | 977 445 | 100 |
| Deprivation score | 38 312 | 97 | 935 128 | 96 |
| CV risk factors | | | | |
| Smoking status | 38 928 | 98 | 669 874 | 69 |
| Blood pressure (systolic) | 38 935 | 98 | 502 188 | 51 |
| BMI (or height and weight) | 38 225 | 97 | 524 384 | 54 |
| Ethnicity | 30 409 | 77 | 585 279 | 60 |
| Cholesterol (HDL, LDL, other) | 37 386 | 95 | 293 082 | 30 |
| Glycaemic control (blood glucose/ HbA1c) | 39 244 | 99 | 253 532 | 26 |

BMI, body mass index; CV, cardiovascular; HDL, high-density lipoprotein; LDL, low-density lipoprotein; RCGP RSC, Royal College of General Practitioners Research and Surveillance Centre.

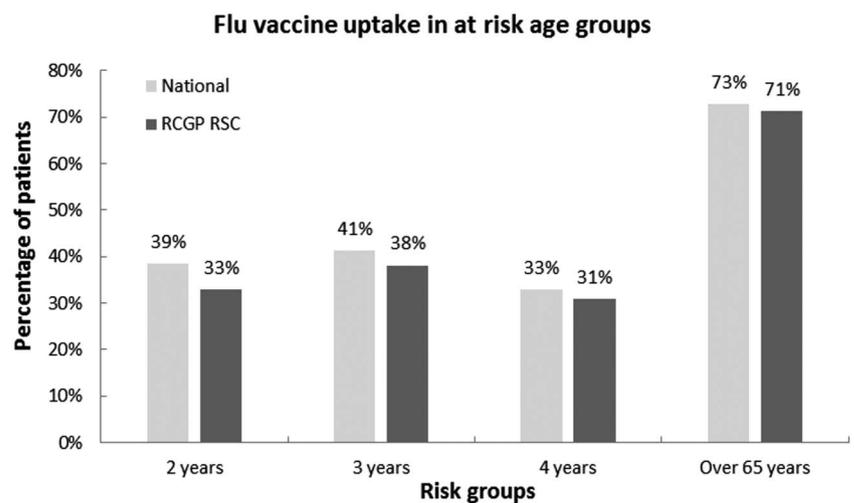**Prescriptions by BNF chapters: National and RCGP RSC**
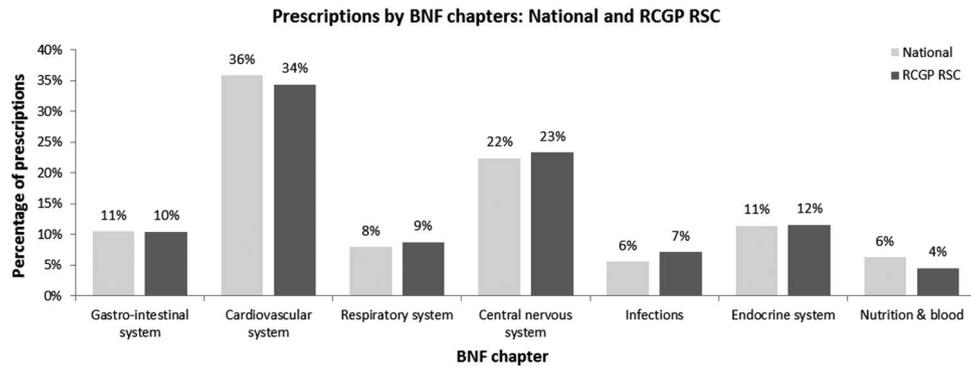


**Figure 4** Prescription items by British National Formulary (BNF) chapter: National and Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC).

areas concerning surveillance, the RCGP RSC database is also representative in a number of clinically relevant areas, which support the network's suitability for real-world evidence research.

This paper provided a clear assessment of the prevalence of chronic conditions when compared with national data, and the completeness of recorded data for risk factors, comparing patients with T2DM and those without. The purpose of this was to assess the suitability of longitudinal research into chronic conditions. It was shown that the data completeness for risk factors drastically improves for patients with an exemplar chronic condition (T2DM), due to the QOF scheme in English primary care.

This paper also highlighted key differences of this network against a national population, due to the nature of its primary purpose. Prescriptions for infections were higher than national levels (though not vaccine uptake), and there was an over-representation of practices in the higher deciles of QOF scores. This is because the RCGP RSC is a sentinel network comprised of volunteer practices, and focused around infectious and respiratory disease surveillance. These differences have been evaluated in this paper, and their relevance can be considered for research using this database.

Most of the differences found were statistically significant; this is due to the large numbers involved in the analysis, which means that very small differences can be easily detected, even when a Bonferroni-Šidák correction is applied to adjust the significance level. Though the difference introduces a bias to the cohort, it has, nonetheless, been quantified in this paper, and can be assessed on a study-by-study basis for its clinical relevance.

## Limitations

An important limitation of this study is that a number of areas (ethnicity, prescriptions, vaccine uptake) were assessed by comparing publically available national rates with data extracted from the RCGP RSC database. The criteria used to extract the network's rates may not have been the same as that used for the published national data, leading to significant differences.

It must be noted that recruitment efforts within the network continue to be expanded, particularly regarding new projects, such as the Integrate study for gastro-intestinal disease surveillance[29] which is focused in the northwest of England; this shift in geographical focus may change the demographic structure of the network. Therefore, the assessment of representativeness presented in this paper may change in future.

**Figure 5** Vaccine exposure: National and Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC).

## CONCLUSIONS AND FUTURE COLLABORATIONS

This network provides a representative sample of the population of England in terms of demographics and clinical outcomes. Future recruitment needs to ensure that, as far as possible, any areas of difference are minimised. The RCGP RSC network, in addition to surveillance, could also be used for research into routine practice, and the interaction of infectious disease with long-term conditions.

The use of large healthcare data sets continues to expand,[30] with many networks having been set up in the UK providing regular updates to large data sets.[31–33] These networks are recognised as having a vital role in disease surveillance, monitoring response to policy change, identification of health-environment interactions and monitoring of side effects of medication.[34–37] More recently, it has been suggested that these large data networks could also be used to perform large-scale clinical trials and ensure real-world medication effectiveness for new medications.[38]

This is particularly important, as clinical trial populations are often not representative of the real-world population, where any interventions or new treatments are ultimately implemented.[39 40] If clinical outcomes are to be measured using these data sets, it is important that they provide a representative sample of the underlying population, reporting any deviations. The RCGP RSC network is representative of the underlying English population, and key significant differences have been clearly quantified in this paper, making this database a rich source for health outcomes research.

It is expected that this database will be available to researchers on a case-by-case basis. Ethical approval by the NHS Research Ethics Committee is needed for data requests to be considered. Data requests for aggregated data will be provided by the University of Surrey team. Researchers wishing to directly analyse the patient-level anonymised data will be required to complete information governance training and work on the data from the secure servers at the University of Surrey. We encourage interested researchers to attend the short courses on how to analyse primary-care data offered by the university twice a year.

Research Ethics Committee is needed for data requests to be considered. Aggregated data tables may be created from the source data to allow specific analyses for approved research and surveillance projects. Researchers wishing to directly analyse the patient-level anonymised data will be required to complete information governance training and work on the data from the secure servers at the University of Surrey. Patient-level data cannot be taken out of the secure servers at the University of Surrey. We encourage interested researchers to attend the short courses on how to analyse primary-care data offered by the university twice a year.

## REFERENCES

1.  Queenan JA, Williamson T, Khan S, et al. Representativeness of patients and providers in the Canadian Primary Care Sentinel Surveillance Network: a cross-sectional study. *CMAJ Open* 2016;4: E28–32.
2.  Fleming DM. Weekly Returns Service of the Royal College of General Practitioners. *Commun Dis Public Health* 1999;2:96–100.
3.  Fleming DM, Schellevis FG, Paget WJ. Health monitoring in sentinel practice networks: the contribution of primary care. *Eur J Public Health* 2003;13(3 Suppl):80–4.
4.  Souty C, Turbelin C, Blanchon T, et al. Improving disease incidence estimates in primary care surveillance systems. *Popul Health Metr* 2014;12:19.
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.  Schlaud M, Brenner MH, Hoopmann M, et al. Approaches to the denominator in practice-based epidemiology: a critical overview. *J Epidemiol Community Health* 1998;52(Suppl 1):13S–19S.
7.  Fleming DM, Zambon M, Bartelds AI. Population estimates of persons presenting to general practitioners with influenza-like illness, 1987–96: a study of the demography of influenza-like illness in sentinel practice networks in England and Wales, and in The Netherlands. *Epidemiol Infect* 2000;124:245–53.
8.  Schwartz E, Kofie VY, Rivo M, et al. Black/White comparisons of deaths preventable by medical intervention: United States and the District of Columbia 1980–1986. *Int J Epidemiol* 1990;19:591–8.
9.  de Lusignan S, Nitsch D, Belsey J, et al. Disparities in testing for renal function in UK primary care: cross-sectional study. *Fam Pract* 2011;28:638–46.
10. Kelleher KJ, Moore CD, Childs GE, et al. Patient race and ethnicity in primary care management of child behavior problems: a report from PROS and ASPN. Pediatric Research in Office Settings. Ambulatory Sentinel Practice Network. *Med Care* 1999;37:1092–104.
11. Clothier H, Turner J, Hampson A, et al. Geographic representativeness for sentinel influenza surveillance: implications for routine surveillance and pandemic preparedness. *Aust N Z J Public Health* 2006;30:337–41.
12. Nutting PA, Baier M, Werner JJ, et al. Practice patterns of family physicians in practice-based research networks: a report from ASPN. Ambulatory Sentinel Pratice Network. *J Am Board Fam Pract* 1999;12:278–84.
13. Pebody R, Warburton F, Andrews N, et al. Effectiveness of seasonal influenza vaccine in preventing laboratory-confirmed influenza in primary care in the United Kingdom: 2014/15 end of season results. *Euro Surveill* 2015;20.
14. Martirosyan L, Arah OA, Haaijer-Ruskamp FM, et al. Methods to identify the target population: implications for prescribing quality indicators. *BMC Health Serv Res* 2010;10:137.
15. Greiver M, Williamson T, Barber D, et al. Prevalence and epidemiology of diabetes in Canadian primary care practices: a report from the Canadian Primary Care Sentinel Surveillance Network. *Can J Diabetes* 2014;38:179–85.
16. Fleming DM, Miles J. The representativeness of sentinel practice networks. *J Public Health (Oxf)* 2010;32:90–6.
17. Kumarapeli P, Stepaniuk R, de Lusignan S, et al. Ethnicity recording in general practice computer systems. *J Public Health (Oxf)* 2006;28:283–7.

18. Office of National Statistics. Postcodes (Enumeration) to output areas to lower layer SOA to middle layer SOA to local authority districts E+W lookup. 2011. https://geoportal.statistics.gov.uk/Docs/Lookups/Postcodes_(Enumeration)_(2011)_to_output_areas_(2011)_to_lower_layer_SOA_(2011)_to_middle_layer_SOA_(2011)_to_local_authority_districts_(2011)_E+W_lookup.zip

19. Department for Communities and Local Government. The English Indices of Deprivation 2015. 2015. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015

20. Office of National Statistics. Clinical commissioning groups (Eng) Jul 2015 Boundaries. 2015. https://geoportal.statistics.gov.uk/Docs/Boundaries/Clinical_commissioning_groups_(Eng)_Jul_2015_Boundaries_(Full_Extent)_V2.zip

21. Bivand R, Lewin-Koh N. maptools: Tools for reading and handling spatial objects. R package version 0.8-27. 2013. http://CRAN.R-project.org/package=maptools

22. Health and Social Care Information Centre. Quality and Outcomes Framework (QOF)—2014–15. 2015. http://www.hscic.gov.uk/catalogue/PUB18887

23. de Lusignan S, Hague N, van Vlymen J, et al. Routinely-collected general practice data are complex, but with systematic processing can be used for quality improvement and research. Inform Prim Care 2006;14:59–66.

24. Croft K; Health and Social Information Centre. Prescription Cost Analysis Data: Prescribing and Primary Care Services. 2015. http://www.hscic.gov.uk/catalogue/PUB17274/pres-cost-anal-eng-2014-rep.pdf

25. Public Health England. Seasonal influenza vaccine uptake amongst GP patients in England: Final monthly data for 1 September 2014 to 31 January 2015. 2015. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/428971/Seasonal_Flu_Jan15_acc2.pdf

26. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995;310:170.

27. Abdi H. The Bonferroni and Šidák corrections for multiple comparisons. Encyclopedia Meas Stat 2007;3:103–7.

28. Kontopantelis E, Reeves D, Valderas JM, et al. Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. BMJ Qual Saf 2013;22:53–64.

29. http://www.integrateproject.org.uk

30. Groves P, Kayyali B, Knott D, et al. The 'big data' revolution in healthcare. McKinsey Quarterly, 2013.

31. Hippisley-Cox J, Stables D, Pringle M. QRESEARCH: a new general practice database for research. Inform Prim Care 2004;12:49–50.

32. Lewis JD, Schinnar R, Bilker WB, et al. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. Pharmacoepidemiol Drug Saf 2007;16:393–401.

33. Bhaskaran K, Forbes HJ, Douglas I, et al. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). BMJ Open 2013;3:e003389.

34. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005;58:323–37.

35. Harcourt SE, Edwards DE, Fleming DM, et al. How representative is the population covered by the RCGP spotter practice scheme? Using Geographical Information Systems to assess. J Public Health (Oxf) 2004;26:88–94.

36. Harcourt S, Smith G, Elliot A, et al. Use of a large general practice syndromic surveillance system to monitor the progress of the influenza A (H1N1) pandemic 2009 in the UK. Epidemiol Infect 2012;140:100–5.

37. Kousoulis AA, Rafi I, de Lusignan S. The CPRD and the RCGP: building on research success by enhancing benefits for patients and practices. Br J Gen Pract 2015;65:54–5.

38. Naci H, Lehman R, Wouters OJ, et al. Rethinking the appraisal and approval of drugs for type 2 diabetes. BMJ 2015;351:h5260.

39. McGovern A, Hinchliffe R, Munro N, et al. Basing approval of drugs for type 2 diabetes on real world outcomes. BMJ 2015;351:h5829.

40. Nallamothu BK, Hayward RA, Bates ER. Beyond the Randomized Clinical Trial: The Role of Effectiveness Studies in Evaluating Cardiovascular Therapies. Circulation 2008;118:1294–303.