How Clinical Practice Research Datalink data are used to support pharmacovigilance

Rebecca E. Ghosh, Elizabeth Crellin, Sue Beatty, Katherine Donegan, Puja Myles and Rachael Williams

Ther Adv Drug Saf
2019, Vol. 10: 1–7
DOI: 10.1177/2042098619854010

Abstract: Pharmacovigilance can be defined as the science of monitoring medicines and vaccines after license for use, the purpose of which is to quantify and characterise the safety profile of a medicine, identify previously unknown adverse reactions, inform risk-benefit assessment, and support the development of actions that can be taken to reduce risks, optimise benefits and monitor their effectiveness. This review discusses the Clinical Practice Research Datalink (CPRD), which is the source of the largest research database in the UK with longitudinal, representative primary care data linked to data from other healthcare settings. CPRD supports international pharmacovigilance by providing a large, anonymised representative general population database with comprehensive capture of patient risk factors and outcomes to researchers within academic, regulatory and pharmaceutical organisations. The specific advantages of CPRD data are discussed in the context of the 'six Vs of big data' including volume, velocity, variety, veracity, validity and value. Examples of where CPRD data have been used for pharmacovigilance research and how these have fed into guidelines and policy are discussed.

Keywords: electronic health records, pharmacovigilance, primary health care, medical record linkage, big data

Received: 30 November 2018; revised manuscript accepted: 18 April 2019.

Correspondence to:
Rachael Williams
Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency, London, UK.
Rachael.Williams@mhra.gov.uk

Rebecca E. Ghosh
Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency, London, UK

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

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

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

Katherine Donegan
Vigilance and Risk Management of Medicines, Medicines and Healthcare Products Regulatory Agency, London, UK

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

Introduction

Pharmacovigilance can be defined as the science of monitoring medicines and vaccines after license for use, the purpose of which is to quantify and characterise the safety profile of a medicine, identify previously unknown adverse reactions, and support the development of actions that can be taken to reduce risks, optimise benefits and monitor their effectiveness. Pharmacovigilance supports effective risk management with the important goal of improving quality of life and safety for patients.[1]

High-quality pharmacovigilance is reliant on high-quality evidence obtained from a variety of sources. Historically, regulatory authorities, such as the UK Medicines and Healthcare products Regulatory Agency (MHRA), have relied on case reports to identify signals from adverse drug reactions.[2] Many of these reports are from voluntary reporting schemes, which may not capture all events, limiting signal identification. Now, however, there is an increasing variety of data sources available which offer scope to expand beyond traditional data collection methods in the type and quality of evidence available for pharmacovigilance.[3] The many types of data sources available for pharmacovigilance research include drug and disease registries, insurance claims databases and electronic health records (EHRs) databases.[3] The increase in population coverage of available EHR data along with increasing numbers of linked datasets means EHR databases can be considered as sources of 'big data'. Marketing authorisation holders have a legal responsibility to collect adverse event reports as well as conduct signal detection and postauthorisation safety studies (PASS).[4] Increasingly, pharmacoepidemiological studies using this wider range of data sources (including EHR databases) are playing a key role in pharmacovigilance activities.[2]

The definition of what constitutes big data varies by context but, from a medical and pharmacovigilance perspective, it has often been described in terms 'six Vs'; data volume, velocity, variety, veracity, validity and value.[5,6] The use of big data EHRs offers new opportunities to generate evidence through the prediction of adverse drug reactions, identification of novel disease and drug interactions and improvements in statistical modelling and simulation methods. Within the European Union, various reviews have found that between one-third and a half of observational PASS submitted to the EMA use EHRs as their primary data source.[7] In the UK, Real World Evidence (RWE) from the CPRD has been used increasingly to inform published treatment guidelines and clinical practice guidance, including from the National Institute for Health and Care Excellence (NICE).[8]

Routinely collected EHRs have several advantages over bespoke data collection, including the speed of access, cost-effectiveness and richness, as well as the size, longitudinal nature and population coverage of the data.[7] However, routine data collection is usually not done with a research focus, leading to data that needs to be cleaned and transformed before it can be repurposed for research. In the UK, there are several sources of primary care data, including The Health Improvement Network (THIN) database,[9] QResearch,[10] the Secure Anonymised Information Linkage (SAIL) Databank in Wales,[11] as well as the Clinical Practice Research Datalink (CPRD).[12] This review discusses the CPRD, which is the source of largest research database in the UK with longitudinal, representative primary care, data linked to data from other healthcare settings.

The CPRD

The CPRD provides some of the largest primary care databases in the world and aims to support international public health research by providing anonymised UK EHRs to researchers within academic, regulatory, and pharmaceutical organisations.[13] CPRD provides primary care data in two combinable databases based on different general practice patient management software system providers: CPRD GOLD based on the Vision® software system and CPRD Aurum based on the EMIS® software system. These two databases have a similar structure and contain anonymised data from general practices who have agreed to provide patient data, with a combined coverage rate of approximately 15% of the UK population.[13] CPRD collects information on demographics, diagnoses, symptoms, signs, prescriptions, referrals, immunisations, behavioural factors and tests. For pharmacovigilance, the indepth prescribing information held in CPRD primary care data is critical, especially in combination with information on population groups that would allow for the identification of specific groups of interest. This includes basic demographics such as age and sex but also allows for the identification of key comorbidities and risk factors such as pregnancy.

The CPRD primary care data are collected during routine general practitioner (GP) care in the UK. As the National Health Service (NHS) in the UK involves the use of GPs as the first point of contact for care, over 98% of the UK population is registered with a GP.[14] This gatekeeper approach makes the GP the first point of contact for many issues and the most likely point at which the first signs or symptoms of drug side effects will be picked up. The UK is also one of the few countries in the world where a patient journey can be followed through EHRs from primary care, secondary care, disease registries and death registries by using the patient's unique patient identifier (NHS Number). CPRD has a deterministic record linkage scheme using the NHS Number which both increases and enhances the available data on patient care, diseases and conditions by linking to additional data sources, including hospitalisation data (Hospital Episode Statistics), death registration data with causes of death (Office for National Statistics death registrations), cancer registrations with data on chemotherapy treatments and radiotherapy (National Cancer Registration and Analysis Service data), mental health and various small area data including the Index of Multiple Deprivation.[15] To date, CPRD has linked data available for 15.9 million unique patients in England with the potential to provide updated linked data every 3 months.[16]

In the following, we describe the specific advantages of the CPRD data with examples of how they have been used for pharmacovigilance, with reference to the main features of big data.

Volume

The size of a dataset can be described both in terms of the number of patients (relating to the power to detect signals), as well as the range of

available data fields and the length of collection time (relating to the range of outcomes and time periods that can be studied). CPRD has been collecting longitudinal data since 1987, and has collated data on over 35 million patient lives covering 15% of the UK population. With 1 in 10 practices currently contributing data to CPRD, the median (IQR) length of available follow-up for CPRD GOLD is 11.97 years (IQR 4.40–23.32) and for CPRD Aurum is 9.13 years (IQR 3.22–20.45) (as of November 2018). Over one-quarter of the patients have over 20 years of follow-up, which is critical for assessing the longer-term effects of drugs. An example of where this has been important is a study comparing the estimated effectiveness of antibiotic classes of respiratory tract infections using data from 1991 to 2012.[17] This study, using data from over two decades and 58 million antibiotic prescriptions, was able to identify the more effective antibiotic classes for respiratory tract infections in the UK.

While analytical techniques such as machine learning and data mining have been in use for decades, the increasingly higher volumes of data available have driven further development in these techniques.[18] Use of such techniques for pharmacovigilance has included text mining and adverse event detection, making use of a wide range of data including social media and EHR databases such as CPRD GOLD.[19–22] Such techniques have the potential to offer more powerful, cost-effective and efficient ways to deal with 'big data' but must still be subject to the same scientific, ethical and governance requirements of all pharmacovigilance research.

CPRD data have been vital for enabling rapid and robust evidence generation as part of a proactive pharmacovigilance programme. An example of this is a study assessing safety data for the multicomponent meningococcal group B vaccine (4CMenB).[23] Prior to this study, the evidence on this vaccine had been limited to clinical trials and localised outbreaks. The UK was the first country to implement a nationwide routine immunisation programme whereby 4CMenB suspected adverse reactions of 4CMenB in children were assessed using the UK Yellow Card Scheme and CPRD data. Between 2015 and 2017, data on 1.29 million children aged 2–18 months receiving 3 million doses of 4CMenB were assessed and it was found that there were similar rates of seizures within 7 days of routine immunisation in the periods before and after 4CMenB introduction. The result was to confirm that, after widespread use of 4CMenB, no significant safety concerns could be identified.[23] Reassuring evidence was also provided that the known reactogenicity of the 4CMenB vaccine had not impacted on the uptake of subsequent doses of other routine vaccinations.

Results from studies using CPRD data have been included in meta-analyses to provide pooled estimates with larger sample sizes, increased precision and broader representativeness to inform international regulatory decision-making. For example, in order to investigate the safety of incretin-based drugs for diabetes care,[24] a common protocol including a nested case-control analysis was used to analyse and then pool CPRD data with those from physician billing claims, hospital discharge abstracts and records of prescription-drug dispensing from four Canadian provinces and the MarketScan database of claims data in the United States. This study found that incretin-based drugs did not increase the risk of hospitalisation for heart failure as compared with commonly used combinations of oral antidiabetic drugs. The combined use of these EHRs allowed for nearly 1.5 million patients to be included in the study, much greater numbers than could have been obtained from clinical trials alone, and for the pooled results to be compared with database specific estimates to explore potential between-site heterogeneity.

Velocity

The more quickly data can be generated and obtained, the more quickly it can be analysed and any safety issues identified and acted upon. CPRD primary care EHR are routinely collected on a daily basis as part of normal clinical care in participating practices,[13] which allows CPRD to release monthly database updates online to approved research groups. This timeliness of data is crucial when a new intervention or medication needs to be investigated for safety, and near real-time vaccine safety surveillance is increasingly used to rapidly detect vaccine safety signals. The feasibility of implementing near real-time vaccine safety surveillance in the monthly CPRD builds has been assessed using seasonal influenza vaccine/Guillain-Barré Syndrome (GBS) and measles, mumps and rubella (MMR) vaccine/seizures as examples.[25] While no specific signals were detected, it was concluded that CPRD data could

be used as a potential data source to monitor the risk of rare outcomes, such as GBS, after seasonal influenza on a near real-time basis.

Another pair of studies on the effects of a pertussis vaccination using CPRD data were influential in shaping vaccination guidelines.[26] In 2012, after several peaks in the UK rates of pertussis infection, the UK's Joint Committee on Vaccination and Immunisation recommended a temporary vaccination programme for pregnant women to protect children against pertussis before they reach their first routine immunisation. It was important to know if this temporary programme offered the anticipated cost-benefit ratio and, more importantly, if there were any unanticipated side-effects. A study using CPRD primary care data on approximately 20,000 vaccinated women identified no increased risks after vaccination for any outcomes studied, including stillbirths and maternal or neonatal death.[26] A concurrently implemented study, combining data on Public Health England's laboratory-confirmed cases and hospital admissions for pertussis in infants between 2008 and 2013, estimated the vaccine effectiveness using estimates of vaccine coverage based on CPRD data.[27] This study confirmed a fall in the number of cases after the introduction of the vaccine programme. Both these studies provided timely evidence on the safety and effectiveness of a programme of vaccination in pregnancy. This subsequently led to the UK Joint Committee on Vaccination and Immunisation (JCVI) recommending pertussis vaccination for all pregnant women.

Variety

Combining data sources not only increases the size of EHRs but also allows for a greater variety of data to be used. The CPRD record linkage scheme has allowed specific pharmacovigilance questions to be addressed that would not be possible using a single source of data. A study using linked CPRD data investigated selective serotonin reuptake inhibitors (SSRIs) use and cancer-specific mortality.[28] This study used CPRD prescription records linked to data from English cancer registries and ONS death registrations (the gold standard source of date and cause of death data in the UK) to create a cohort of 23,669 patients with newly diagnosed breast cancer between 1998 and 2012. SSRI use was found to be associated with a 27% increase in breast cancer mortality, an association that may be partly explained by confounding by indication, as the increased risk was attenuated when restricting to patients with prior depression or other depression medication.

Additional variety is added to the CPRD data through added products such as the mother–baby link, which is an algorithm to identify pairs of mothers and babies in the primary care data, enabling the longitudinal EHR data from both mother and child to be combined,[29] and the pregnancy register, a more complex algorithm for identifying the start, end and outcomes of pregnancies as recorded in primary care.[30] The mother–baby link was used in the pertussis vaccination study to look at the health of both the mother and the baby.[26]

Veracity

The veracity of a data source is determined by both the data source and how the data are processed. The CPRD data are broadly representative of the UK population in terms of age, sex and ethnicity.[13] The uncertainties related to the use of 'big data' resources, including how best to combine and use these complex resources in the most effective way, remains a challenge for pharmacovigilance. Evidence generated by EHRs is often used to inform important pharmacovigilance decisions and this evidence is generated from resources that were not specifically designed for research but are a by-product of a complex healthcare system.[31]

The quality of primary care data in the UK is partly driven by the Quality and Outcomes Framework (QOF), which is a voluntary annual reward and incentive programme for GP surgeries. The QOF scheme, introduced in 2004, promotes the recording of key data items such as smoking status and ethnicity by GPs and has been shown to improve the quality of key data items.[32] CPRD internal processes also assure researchers by conducting approximately 1300 data quality and validation checks and by providing sets of data quality criteria for the primary care data. The main quality criteria of use to pharmacovigilance research include an 'acceptability for research' flag for patients, which is based on registration status, recording of events in the patient

record, and valid age and gender. A practice level data quality metric, 'up to standard' (UTS) time is calculated based on continuity of recording and is calculated for each practice from the date at which it meets minimum quality criteria. These data quality flags can be used by researchers to select research-quality patients and periods of quality data recording to ensure the veracity of the data for pharmacovigilance research.[13] The quality of the data is crucial for research, especially when dealing with controversial issues. Therefore, in addition to CPRD checks and flags, researchers are encouraged to undertake their own quality checks before use of the data to explore and account for variations in data entry, including missing data, across patients, GP practices and calendar time.[13]

An example where high-quality evidence is critical is the worldwide scare over a potential link between the MMR vaccine and autism, started by a 1998 paper published in the Lancet, subsequently retracted due to false claims that patients were 'consecutively referred' and that the investigations were 'approved' by the local ethics committee.[33] A case-control study using data from CPRD (then the UK General Practice Research Database – GPRD) was integral to demonstrating the lack of association with the risk of pervasive developmental disorders.[34] Cases were patients born in 1973 or later with a recorded diagnosis of pervasive developmental disorder and were matched to controls on age, sex and the general practice they were registered with. The study found 78% cases had MMR vaccine recorded before diagnosis, compared with 82% controls before the age at which their matched case was diagnosed. This evidence was used by The National Institute for Health and Care Excellence (NICE) to develop clinical and drug safety guidelines and has been crucial in helping to restore medical and public opinion.

Value

The value of any big data for pharmacovigilance is whether the costs and benefits of collecting and analysing the data are balanced in a way that keeps the resource available for research. In addition to the quality checks and indicators that CPRD provides, value is added to the data through the record linkage programme and the additional products such as the mother–baby link[29] and the pregnancy register.[30] In addition, CPRD provides cohort identification and data extraction tools to researchers as well as data and coding dictionaries.

CPRD also offers supplemental data collection services such as GP and patient questionnaires to augment EHR-based research and validate findings. GP questionnaires can be used to validate both the exposure and the outcome while patient questionnaires can add data not recorded by GPs, such as symptom diaries. A recent study supplemented CPRD data with saliva sample collection and patient drug-use diaries to investigate adrenal insufficiency following glucocorticoid exposure in patients with rheumatoid arthritis.[35] This study successfully collected supplemental data from patients and in doing so highlighted some important considerations for future studies, including the need for engagement with patients and GPs to maximise recruitment rates. There is also the potential to support and conduct EHR-enabled pragmatic clinical trials starting with trial feasibility and protocol optimisation using near real-time estimates on eligible patient pools, through to locating eligible patients in primary care for recruitment or referral. These additional research services provided by CPRD can aid pharmacovigilance research, helping to supplement data and validate findings for observational research as well as through improving the efficiency and cost-effectiveness of clinical trials.

Conclusions

Pharmacovigilance and risk management have aimed to move from a reactionary framework towards more proactive ongoing monitoring of the benefits and risks of medications and interventions; actively seeking data on the safety of medicines and vaccines, as well as just trying to identify new risks, and monitoring the effectiveness of risk minimisation measures so that further changes in the use and safety profile of a medicine can be identified and any further action needed can be more rapidly implemented. From a regulatory perspective, authorisation of medicine and products has traditionally been based on high-quality data of a well-known provenance such as randomised clinical trials. Use of big data can provide evidence more quickly and cost-effectively but data provenance is key. Databases such as CPRD offer volume, velocity, variety and

veracity while also offering additional research services to supplement EHR data. This enables CPRD to support international pharmacovigilance by providing a large, anonymised representative general population database with comprehensive capture of patient risk factors and outcomes to researchers within academic, regulatory, and pharmaceutical organisations.

Funding

CPRD is jointly sponsored by the UK government's Medicines and Healthcare products Regulatory Agency and the National Institute for Health Research (NIHR). As a not-for-profit UK government body, CPRD seeks to recoup the cost of delivering its research services to academic, industry and government researchers through research user licence fees.

Conflict of interest statement

All authors are employed by the MHRA, which is an Executive Agency of the Department of Health, but have no conflicts of interest that are directly relevant to the contents of this study. The MHRA has statutory responsibility for the pharmacovigilance of medicinal products on the UK market.

ORCID iD

Rebecca E. Ghosh https://orcid.org/0000-0001-6009-3040

References

1. Fornasier G, Francescon S, Leone R, *et al*. An historical overview over Pharmacovigilance. *Int J Clin Pharm* 2018; 40: 744–747.
2. Lane S, Lynn E, Shakir S, *et al*. Investigation assessing the publicly available evidence supporting postmarketing withdrawals, revocations and suspensions of marketing authorisations in the EU since 2012. *BMJ Open* 2018; 8: 19759.
3. Bate A, Reynolds RF and Caubel P. The hope, hype and reality of Big Data for pharmacovigilance. *Ther Adv Drug Saf* 2018; 9: 5–11.
4. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) - Module VIII - Post authorisation safety studies (Rev 3), https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-viii-post-authorisation-safety-studies-rev-3_en.pdf (2017, accessed 22 May 2019).
5. Khan MAUD, Uddin MF and Gupta N. Seven V's of Big Data understanding Big Data to extract value. In: *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education – 'Engineering Education: Industry Involvement and Interdisciplinary Trends', ASEE Zone 1 2014*, Bridgeport, CT, USA, 3–5 April 2014, 14334154, pp. 1–5. Piscataway, NJ: IEEE.
6. Trifirò G, Sultana J and Bate A. From big data to smart data for pharmacovigilance: the role of healthcare databases and other emerging sources. *Drug Saf* 2018; 41: 143–149.
7. Pacurariu A, Plueschke K, McGettigan P, *et al*. Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation. *BMJ Open* 2018; 8: 23090.
8. Oyinlola JO, Campbell J and Kousoulis AA. Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidances. *BMC Health Serv Res* 2016; 16: 299.
9. The Health Improvement Network (THIN). www.visionhealth.co.uk/portfolio-items/-the-health-improvement-network-thin/ (2018, accessed 22 May 2019).
10. University of Nottingham. QResearch. https://www.qresearch.org/ (2018, accessed 22 May 2019).
11. SAIL Databank. SAIL Databank - The Secure Anonymised Information Linkage Databank. https://saildatabank.com/ (2019, accessed 22 May 2019).
12. CPRD. Clinical Practice Research Datalink - CPRD. https://www.cprd.com/ (2018, accessed 22 May 2019).
13. Herrett E, Gallagher AM, Bhaskaran K, *et al*. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; 44: 827–836.
14. NHS Digital. Attribution Data Set GP-Registered Populations Scaled to ONS Population Estimates - 2011. https://digital.nhs.uk/data-and-information/publications/statistical/attribution-dataset-gp-registered-populations/attribution-data-set-gp-registered-populations-scaled-to-ons-population-estimates-2011 (2012, accessed 22 May 2019).
15. Padmanabhan S, Carty L, Cameron E, *et al*. Approach to record linkage of primary care data

from Clinical Practice Research Datalink to other health-related patient data: overview and implications. *Eur J Epidemiol* 2018; 34: 91–99.

16. Ghosh RE, Padmanabhan S, Williams R, *et al*. Including primary care data from multiple software systems in a data linkage programme: results from expanding the Clinical Practice Research Datalink (CPRD). In: *Abstracts of the 34th International Conference on Pharmacoepidemiology & Therapeutic Risk Management,* Prague Congress Centre, Prague, Czech Republic, 22–26 August 2018, p. 89. New York: Wiley.
17. Berni E, Butler CC, Jenkins-Jones S, *et al*. Comparative estimated effectiveness of antibiotic classes as initial and secondary treatments of respiratory tract infections: longitudinal analysis of routine data from UK primary care 1991–2012. *Curr Med Res Opin* 2016; 32: 1023–1032.
18. Zhang L, Tan J, Han D, *et al*. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov Today* 2017; 22: 1680–1685.
19. Ben Abacha A, Chowdhury MFM, Karanasiou A, *et al*. Text mining for pharmacovigilance: using machine learning for drug name recognition and drug–drug interaction extraction and classification. *J Biomed Inform* 2015; 58: 122–132.
20. Wang W, Haerian K, Salmasian H, *et al*. A drug-adverse event extraction algorithm to support pharmacovigilance knowledge mining from PubMed citations. AMIA. *Annu Symp proceedings AMIA Symp* 2011; 2011: 1464–1470.
21. Sloane R, Osanlou O, Lewis D, *et al*. Social media and pharmacovigilance: a review of the opportunities and challenges. *Br J Clin Pharmacol* 2015; 80: 910–920.
22. Weng SF, Reps J, Kai J, *et al*. Can machine-learning improve cardiovascular risk prediction using routine clinical data? (ed. B Liu). *PLoS One* 2017; 12: e0174944.
23. Bryan P, Seabroke S, Wong J, *et al*. Safety of multicomponent meningococcal group B vaccine (4CMenB) in routine infant immunisation in the UK: a prospective surveillance study. *Lancet Child Adolesc Heal* 2018; 2: 395–403.
24. Filion KB, Azoulay L, Platt RW, *et al*. A multicenter observational study of incretin-based drugs and heart failure. *N Engl J Med* 2016; 374: 1145–1154.
25. Leite A, Thomas SL and Andrews NJ. Implementing near real-time vaccine safety surveillance using the Clinical Practice Research Datalink (CPRD). *Vaccine* 2017; 35: 6885–6892.
26. Donegan K, King B and Bryan P. Safety of pertussis vaccination in pregnant women in UK: observational study. *BMJ* 2014; 349: g4219.
27. Amirthalingam G, Andrews N, Campbell H, *et al*. Effectiveness of maternal pertussis vaccination in England: an observational study. *Lancet* 2014; 384: 1521–1528.
28. Busby J, Mills K, Zhang SD, *et al*. Selective serotonin reuptake inhibitor use and breast cancer survival: a population-based cohort study. *Breast Cancer Res* 2018; 20: 4.
29. Boggon R, Gallagher A, Williams TJ, *et al*. Creating a mother baby link and pregnancy register for a UK population. *Pharmacoepidemiol Drug Saf* 2011; 20: 44–45.
30. Campbell J, Williams R, Minassian C, *et al*. Identifying and validating pregnancy episodes in primary care electronic health records. In: *SAPC Society for Academic Primary Care, Annual Society Meeting*. London, 10–12 July 2018; London: Society for Academic Primary Care.
31. Hall GC, Sauer B, Bourke A, *et al*. Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2012; 21: 1–10.
32. Doran T, Kontopantelis E, Valderas JM, *et al*. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ* 2011; 342: d3590.
33. Wakefield AJ, Murch SH, Anthony A, *et al*. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet (London, England)* 1998; 351: 637–641.
34. Smeeth L, Cook C, Fombonne E, *et al*. MMR vaccination and pervasive developmental disorders: a case-control study. *Lancet (London, England) 2004*; 364: 963–969.
35. Joseph RM, Soames J, Wright M, *et al*. Supplementing electronic health records through sample collection and patient diaries: a study set within a primary care research database. *Pharmacoepidemiol Drug Saf* 2018; 27: 239–242.