Clinical epidemiological studies of drug safety and disease risk factors using large primary care databases

By Yana Vinogradova

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy by

published works.

2017

Abstract

Background

Observational studies of drug treatments complement pre-marketing drug trials and provide real-world outcomes of effectiveness and safety. Large UK primary care databases offer cost-effective access to clinical information for long-term studies requiring great statistical power and deliver findings representative of the general population. However, such data are not collected primarily for research, so all share weaknesses that must be offset by sophisticated use of statistical methodologies. This paper clarifies the current strengths and limitations of these data sources and discusses their potential. In the context of routinely collected primary care data sources, studies focusing on drug safety are used to show appropriate application of statistical techniques, and present a contribution to existing methodological practice based on multi-database use.

Methods

Methodological approaches to address potential data-related biases and drug safety study-related issues are discussed. These include coding differences, analyses of exposure, confounding factors to be included in models, missing data, misclassification bias from outcome uncertainty and prescription-only information, and use of sensitivity analyses to estimate the impact of information gaps and verify the validity of findings. A novel application of triangulation between the findings of separate identical analyses of two databases is introduced.

Results

The submitted papers are used to exemplify the delivery of more accurate estimates of risks than previous studies, with further comment on how the methodologies were used

to address potential issues. The results of triangulation between the findings of two separate identical analyses based on different databases show confidence intervals for the combined results on average 30 per cent narrower than those of the original analyses.

Conclusions

The application of emerging/developing methodologies enables large UK primary care databases with national coverage to deliver robust findings applicable to the general population and derived from long-term studies with great statistical power. The potential for future development is also shown, including use of multi-databases to further increase statistical power.

Acknowledgements

I would like to acknowledge the support, advice and encouragement of my advisor, Professor Carol Coupland.

I am also very grateful to my line manager, Professor Julia Hippisley-Cox, for her support and advice in the creation of this text, and in particular for her encouragement of my development into a fully-fledged researcher, providing me with opportunities to extend my skills in a wide range of projects and supporting my role as lead investigator.

I acknowledge the contribution of EMIS and the University of Nottingham for expertise in creating and maintaining QResearch® and of the EMIS practices that contributed data. I thank CPRD and Vision Practices for allowing access to the CPRD for my studies.

I would also like to thank Professor Joe Kai for his encouragement, and for his advice prior to submission of this text.

List of Published works

1. Vinogradova Y, Hippisley-Cox J, Coupland C, Logan R. Risk of colorectal cancer in patients prescribed statins, nonsteroidal anti-inflammatory drugs, and cyclooxygenase-2 inhibitors: nested case-control study. *Gastroenterology*. 2007;133:393-402.
2. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to statins and risk of common cancers: a series of nested case-control studies. *BMC Cancer*. 2011;11:409.
3. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to cyclo-oxygenase-2 inhibitors and risk of cancer: nested case-control studies. *British Journal of Cancer*. 2011;105:452-9.
4. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of cancer: a protocol for nested case-control studies using the QResearch primary care database. *BMJ Open*. 2012;2(1):e000548.
5. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of gastrointestinal cancers: series of nested case-control studies with QResearch and CPRD data. *BMJ*. 2013;346:f114.
6. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of non-gastrointestinal cancers: series of nested case-control studies using QResearch and CPRD data. *British Journal of Cancer*. 2013;109:795-806.
7. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to combined oral contraceptives and risk of venous thromboembolism: a protocol for nested case-control studies using the QResearch and the CPRD databases. *BMJ Open* 2014;4(4):e004499.

8. Vinogradova Y, Coupland C, Hippisley-Cox J. Combined oral contraceptives and risk of venous thromboembolism: nested case-control studies using the QResearch and the CPRD databases. *BMJ*. 2015;350:h2135.
9. Hippisley-Cox J, Vinogradova Y, Coupland C, Parker C. Risk of malignancy in patients with schizophrenia or bipolar disorder: nested case-control study. *Archives of General Psychiatry*. 2007;64(12):1368-76.
10. Vinogradova Y, Coupland C, Hippisley-Cox J, Whyte S, Penny C. Effects of severe mental illness on survival of people with diabetes. *British Journal of Psychiatry*. 2010;197(4):272-7.
11. Vinogradova Y, Hippisley-Cox J, Coupland C. Identification of new risk factors for pneumonia: population based case control study. *British Journal of General Practice*. 2009;59(567):742-9.
12. Vinogradova Y, Coupland C, Hippisley-Cox J. Risk of pneumonia in patients taking statins: population-based nested case-control study. *British Journal of General Practice*. 2011;61(592):e742-e8(7).

Contents

Extended abstract

Background

Introduction

An important aspect of pharmacoepidemiology is the risks posed by medications. Randomised controlled trials, conducted prior to marketing, test for safety and common side-effects. These cannot reliably assess real-world, long-term risks in the general population because of cohort selectivity, close management and relatively short-term follow-ups. Risks of uncommon or latent adverse effects in the general population can be assessed only through observational studies designed to overcome these weaknesses. Sometimes, the rarity of adverse conditions or a desire to compare details of different exposures may require extremely large cohorts with long periods of follow-up, restricting sources of data to the largest available. For such studies, electronic healthcare data, routinely collected in many parts of the world, offer valuable material.

Around the world, these include: administrative data linked to prescription information, physician services use, hospital discharges and vital statistics in Denmark[1,2], Canada[3], the Netherlands[4] and Italy[5]; health insurance and healthcare plan databases in Finland[6], Taiwan[7], the USA[8] and Israel[9]; and general practice databases in New Zealand[10] and Spain[11]. However, the source and purpose of such data collections are various and must be considered. General practice databases, prospectively recording information from treatment centres, are clearly likely to be among those most suitable for the assessment of real-world, long-term outcomes in the general population. A number of such databases exist within the UK, the two largest of which are QResearch® (www.QResearch.org) and Clinical Practice Research Datalink (CPRD, www.cprd.com), which have been the basis for all the research submitted in this thesis.

Aims and objectives

The aim of each submitted study is to contribute to the epidemiology of either drug safety or disease risk factors. Experience gained from protocol development, data preparation, analysis, writing and defence of the submitted studies in challenging publication review processes is used to address the place of large primary care data sources in healthcare studies. Information from further research in the preparation of this text has also been incorporated to assess their current strengths and limitations, and their potential as evidence sources for reliable assessment of real-world outcomes in the general population.

In the context of routinely collected primary care data sources, the aim of this extended summary is to demonstrate appropriate use of statistical methodologies, including a novel use of meta-analytic techniques to facilitate multi-database design to overcome the limitations of low power for studies of rare conditions and long-term treatments. The objectives are to describe and critically discuss the studies' contributions to existing methodology, to provide recommendations for researchers based on expertise and experience gained, to comment on anticipated and desirable changes in database content, and to discuss possible future healthcare database development and strategy.

Healthcare data collection in the UK

Primary care data in the UK have been recorded electronically since the 1990s. The NHS has encouraged the use of practice-based software and the development of electronic patient records[12], while some research teams have been given specific NHS research funding to stimulate and develop primary care research[13,14]. This has facilitated computerisation of administrative tasks, such as disease registers and prescription recording, and of medical records generated by doctors, nurses and administrators.

Collecting and encoding primary care information in the UK

Three principal information systems are used in UK general practices to encode medical and administrative information. Two of these – Egton Medical Information Services (EMIS) Web, used by 55 per cent of UK practices (www.emis-online.com), and INPS Vision from In Practice Systems Ltd, used by 10 per cent of UK practices (www.inps.co.uk/vision) – employ the Read Version 2 encoding system. The third, TPP SystmOne, which is used by 35 per cent of practices, uses Read Version 3, usually known as Clinical Terms Version 3 (CTV3) to signify that it is a completely different encoding system although also based on Read codes. Read coding is the standard clinical terminology system used in UK general practice. The codes are published under Crown Copyright: they were mandated by the NHS in April 1999 and their intellectual property was purchased by the UK government.

Read Version 2 is a five-level hierarchical system for encoding medical and non-medical terms, including diagnoses, symptoms, patient occupations, medical procedures generally used in UK primary care, and administrative processes[15]. The Read system contains over 150,000 codes, many of which are related in a complex fashion to a single morbidity. However, existing codes do not cover all possible situations, so some information has to be entered as free text[16] which, for confidentiality reasons, is usually unavailable to researchers. Dosage information, which *is* available, is also initially recorded as free text, although in some databases it is subsequently encoded. CTV3 incorporates a radically different approach to the relationship between codes, replacing the hierarchy within codes themselves with an external table of binary hierarchical code relationships to deliver hierarchies of any depth and other advantages.

Most suppliers of encoding systems provide their own code search engines, which aim to optimise the search process for clinicians and other practice staff[17]. All incorporate

different architectures and technical solutions and employ different mechanisms for code retrieval. To improve healthcare or better manage chronic conditions, a practice or network of practices may create bespoke templates to facilitate the process of code selection in a particular area or collect additional information[18]. This may create differences in code selection or recording quality between practices. Whether or not templates are used, the different code identification strategies for encoding systems and the order in which underlying codes are revealed – whether in alphabetical order, or starting with those most regularly used – may result in the selection of different Read codes for the same condition. This is particularly an issue where an unusual condition or an unusual level of detail is required.

Linkages to external data

Not all relevant information is recorded in primary care electronic health records, which may hinder research on rare outcomes. Hospitalised patients, and those who were diagnosed post-mortem or have since died, may have no relevant records held by the practice, so it is important also to access available complementary sources to maximise data completeness. Hospital Episode Statistics (HES) (http://www.hscic.gov.uk/hes) is a data repository of all admissions, outpatient appointments and A&E attendances at NHS hospitals in England. Data are coded using the International Statistical Classification of Diseases and Related Health Problems (ICD-10). Another current source of linkage data is Cancer Registry UK, which provides details of cancer diagnoses in the UK (http://www.ncin.org.uk/). The Office for National Statistics (ONS) holds mortality data across the UK, including causes, again using ICD-10.

Information on socio-economic status is often not recorded by general practices but is important to researchers. It can be obtained from 2001/2011 UK census data in the form of area-level (postcode-derived) Townsend composite scores based on data for

unemployment, overcrowding, and home and car ownership[19] and available at patient or practice level using residence or practice postcodes. As this may lead to patient identification in thinly populated areas, Townsend scores are commonly supplied in quintiles or deciles. Higher scores indicate greater deprivation, associated with higher risks of cancer[20], pneumonia[21] and mortality[22].

Ethnicity may reflect both lifestyle factors and genetic tendencies. It is self-identified but not always recorded for every patient within a general practice. HES provides another source of ethnicity data, enabling some missing data to be accessed for practices with this linkage.

Primary care databases

General practice primary care databases of various sizes and scope exist in the UK. Some of these focus on a geographical region or a specific condition or area of research. Lower geographic coverage or size may be offset by higher levels of completeness and accuracy through closer data management, sometimes including supplementary information from patients or clinicians. Use of such databases is likely to be limited to local or regional users unless a data aggregation agent such as PRIMIS (http://www.nottingham.ac.uk/primis/) is used. However, the largest national databases aim for coverage as closely representative of the national population as possible and are generally available for research. QResearch® and CPRD are the two largest, and are available to and used by researchers worldwide. Other well-known databases include:

- The Health Improvement Network (THIN), a national database using Vision software for data collection. This is managed by IMS Health (www.csdmruk.imshealth.com), a company providing access to the data for use in medical research, with over 550 practices. Data are collected prospectively, and a

number of contributing practices are also linked to HES, but researchers must be aware that many practices contribute to both THIN and CPRD, with the precise proportion being uncertain.

- Launched in 2013, a new national database, ResearchOne, has been developed by the TPP IT company (provider of SystmOne) in partnership with the University of Leeds (www.researchone.org/). This contains patient record information for over four million people collected from more than 300 health and social care organisations. TPP has a strong presence in Yorkshire and the Humber, so is most representative of this region.
- The Consultations in Primary Care Archive (CiPCA) (www.keele.ac.uk/mrr/cipcadatabase/) is a database established by Keele University and North Staffordshire Health Authority to collect data from local practices. All practices follow established procedures for data audit, programmes of training on morbidity coding and validation[23], so data quality is claimed to be at least as good as larger general practice databases. Research projects using it are focusing on musculoskeletal disorders and related pain issues, so completeness and quality are likely to be highest in data relating to these.
- Another example of a research-focused database is The Optimum Patient Care (OPC) Research Database (http://optimumpatientcare.org/opcrd/). It provides data for respiratory research and offers a free respiratory review service. OPC contains data from over 550 general practices in the UK. The practices use various IT systems and the data are collected with specially developed software. This database also contains both longitudinal medical records and questionnaire-based information on patient-reported outcomes[24].

CPRD

CPRD, with more than 600 contributing practices, is older than QResearch®. Its origins lie in a database established in 1987 by an Essex general practitioner to share information between practices. In 1993, it was taken over by Reuters Health Information, which in 1994 donated it to the Department of Health, when it was renamed the General Practice Research Database (GPRD). In 2012, GPRD became part of the CPRD research service jointly funded by the NHS National Institute for Health Research and the Medicines and Healthcare Products Regulatory Agency. Historically, CPRD has used INPS Vision to capture general practice information, but is now also incorporating information from practices using EMIS Web.

GPRD/CPRD has been extensively validated, with practices reviewed for consistency of information[25]. Data quality is maintained by the MHRA research team[26] and its value has been demonstrated in more than a thousand peer-reviewed publications since 1994. A systematic review based on 46 such studies[27], focusing on the accuracy and completeness of diagnostic coding, demonstrated close agreement between disease prevalence rates in GPRD/CPRD and other sources of information, including HES, the British National Household Survey, the Doctors' Independent Network[28], and the MediPlus primary care database[29].

QResearch®

QResearch® is currently the largest general practice database in the UK, with over 1,300 practices, more than 23 million patients and 25 years of clinical records. It derives from a project in 2002 between Egton Medical Information Services (EMIS) and the University of Nottingham to collect high-quality primary care data for medical research[30], and all contributing practices use EMIS Web for encoding general practice

information. To encourage continuous development, practices contributing to QResearch® receive feedback on the quality of their data[30].

QResearch® was extensively validated early on. Demographic measures, such as age–sex distributions and birth/death rates, were compared with census data and figures from the ONS; consultation rates and prevalence rates for common conditions with data from the General Household Survey, GPRD and other databases; and prescribing rates with prescribing analysis and cost data from the National Institute for Health and Care Excellence (NICE). Close correspondence of profiles with these sources was revealed, together with high levels of completeness and consistency[31]. The quality and value of QResearch® data have since been demonstrated in over two hundred peer-reviewed publications.

Similarities and differences between QResearch® and CPRD

Both databases are not-for-profit research services, and doctors in participating practices receive no particular training. Both contain information gathered prospectively over long periods on patient registration details, sex, year of birth, ethnicity, smoking habits and alcohol consumption. General practitioners record details of illnesses, new symptoms and diagnoses, and family history. Details of all further clinical contacts, such as laboratory tests, referrals to specialists and hospitalisations are then added. Prescription records are well-documented, being recorded at the time of prescribing. Patient records enter the database from their time of registration with a participating practice, leaving only if they transfer to a non-participating practice, quit the NHS or die.

The Dictionary of Medicines and Devices (http://dmd.medicines.org.uk/) is used by both INPS and EMIS, but prescribing information is collected using different software

– Gemscript in INPS (CPRD) and E-prescribing in EMIS (QResearch®). Dosage instructions are recorded in free text, but in CPRD 95 per cent of the most frequent dosages are coded by the CPRD Operations Team before release to researchers, while in QResearch® original text is available. For CPRD, the MHRA multi-disciplinary team that encodes prescribing also reviews data quality by checking consistency and linking to other sources.

For both databases, data are pseudonymised. For QResearch®, all data are anonymised at source according to the ICO code on anonymisation[32], so no strong patient identifiers are taken from practices: dates of birth are changed to years of birth, postcodes mapped to deprivation scores with only the deprivation scores taken, and NHS numbers pseudonymised using a non-reversible hashing algorithm. For CPRD, however, fully identified data are taken from general practices, including NHS number, full birth date and postcode, and the data are then de-identified by NHS Digital before being passed to CPRD. Practices contributing to CPRD must therefore inform patients that they are passing on data prior to pseudonymisation and must ensure that they respect patient objections.

The data are then uploaded to dedicated servers. QResearch® data are transmitted to the University of Nottingham, where QResearch® is linked to other data sources and made available to researchers. All QResearch® practices and the majority of CPRD practices are linked to Cancer Registry data and ONS Mortality data. English practices with these linkages (100 per cent in QResearch® and 75 per cent in CPRD) are also linked to HES data, coverage of which is however limited to England. QResearch® data access is covered by ethical approval from Derby Research Ethics Committee once a protocol has been approved by the QResearch® Scientific Committee. CPRD access requires

approval from the Independent Scientific Advisory Committee for MHRA Database Research.

Recent direct comparisons between QResearch® and CPRD have shown high levels of similarity and have concluded that the databases are likely to give similar results in epidemiological studies[33,34]. Age and sex distributions and recording levels for clinical outcomes and clinical values are similar for both. Differences include the proportion of available patient-level deprivation data (as measured by Townsend scores), with full coverage in QResearch® and partial coverage in CPRD (58 per cent). In CPRD, each patient is allocated to the relevant decile of Townsend score, whereas in QResearch® individual Townsend scores are also available on site at Nottingham. Ethnicity is better recorded in QResearch® (58 versus 32 per cent). Practices contributing to QResearch® and CPRD are spread throughout the UK, with the majority in England. Coverage within English geographical areas differs in detail, but both databases reflect a higher population density toward the south east[33].

Using QResearch® and CPRD for research

Advantages

Both databases are representative of the general population, so results from either are more generalisable than those from studies using recruitment. Routinely collected information is not limited by specific aims, and there are unlikely to be any systematic differences between various groups of patients with regard to accuracy or completeness of their records. Unlike questionnaire-based studies, where only living and consenting patients can be included, individual consent is not required for use of anonymised data in ethically approved projects, so data for all relevant patients can be included.

Comprehensive prescribing details, including duration and dose, have also been gathered for long periods, facilitating both assessment of biological gradients for exposures and investigation of different or emerging drug types for comparative associations of outcomes. The databases contain information relating to important confounders – comorbidities, other medications and lifestyle measures – so much of the data required to increase accuracy and reliability of results is available within each source.

While there are differences between QResearch® and CPRD in how data are collected and managed, these are minor when compared to their similarities, making them potentially interchangeable as research resources. The most recent of the submitted projects demonstrate this in the application of a meta-analytic technique to triangulate between the results of separate identical studies based on QResearch® and CPRD. This facilitates more detailed investigation and delivers more accurate and robust estimates, and has potential importance especially for research into very rare conditions or to maximise the level of detail possible in investigations including dosages and/or different drug types.

Limitations

Despite the quantity and quality of data, both databases have limitations that must be considered. Lifestyle information, such as smoking and body mass index (BMI), may be relevant to some studies, but historically was not consistently recorded. BMI recording in CPRD increased from 37 per cent for 1990–1994 to 77 per cent for 2005–2011, but only 53 per cent of patients were found to have a recent (within three years) record of BMI[35]. Smoking was recorded for 81 per cent of patients in CPRD between 2007 and 2011[36]. Investigation of CPRD data has similarly shown that missing results

for blood pressure tests are not completely random, which is logical because patients with higher blood pressure tend to have more records[37].

The databases also lack completeness in formally adjudicated outcomes. For patients referred to secondary care, subsequent external prescribing and diagnostic information are not always available, and potentially useful information in correspondence relating to external consultations is in free-text form, so access to it is very limited and expensive because of confidentiality issues. A systematic review based on 212 studies investigated the validity of 183 different diagnoses in GPRD/CPRD and reported that the median proportion of cases with diagnosis confirmed by internal and external data was 89 per cent, with varying proportions for different disease groups – 95 per cent for cancer but only 57 per cent for blood-related illnesses[38]. Use of over-the-counter medications and data such as occupation, lifestyle, diet and involvement in sport are also not always available.

Observational studies – gaps and further improvements

Traditional observational studies have often been too short to investigate the effects of long-term prescribing, and many have been prone to a range of data collection biases such as selection or recall. Studies using large, routinely-collected primary care data sources have avoided these data-related issues but, regardless of data source used, some previous studies, including quite recent ones, have failed adequately to address issues such as duration of drug exposure[39] and adjustment for important confounders[40]. Some have also simply excluded observations with missing data or treated missing values as a separate category[41]. Although QResearch® and CPRD facilitate long-term, comprehensive and detailed drug safety research, challenges remain because the data consist of unmediated clinical records. These include inconsistencies in coding, large numbers of missing values for some measures and a lack of some potentially important

information for certain studies. In the discussion below, some developmental improvements are suggested, which might reduce information gaps and improve the utility of these databases.

Methods

While large numbers of participants and long periods of collection may offer a foundation for inherently powerful studies and reliable risk estimates, realising this potential requires close attention to detail. Biases relating to collection and selection issues, missing information, and confounding caused by inadequate study design may all lead to spurious conclusions. The main challenge for researchers is to ensure the accuracy of risk assessments by taking into account the specific characteristics of any dataset selected as a research base and making appropriate use of all methodological tools developed to overcome problems with datasets.

Included studies

Most of the submitted research has consisted of studies of associations between commonly used drugs and risks of cancers. Paper 1 concentrates on statins and colorectal cancer risks, while Papers 2 to 6 all consider risks for the ten most common cancer sites – Paper 2 on associations with statins, Paper 3 on associations with cyclooxygenase-2 (COX2) inhibitors, and Papers 4 (protocol), 5 and 6 on associations with bisphosphonates. Papers 7 (protocol) and 8 consider use of combined oral contraceptives and venous thromboembolism (VTE) risk. Papers 5, 6 and 8 describe studies that used the combined results of separate but identical QResearch® and CPRD analyses to enhance the accuracy of association estimates. Papers 9 to 12 report on four earlier studies, which indicate early development of this research activity, demonstrating both, how the studies were improved by incorporating emerging methodological developments and the potential for use of large-scale primary care databases in other research areas.

Each submitted study has its merits, but as understanding of the databases and experience of using them has increased, and as relevant statistical methodologies have emerged or improved, so the sophistication of the approach has developed. Therefore, the primary focus of this extended abstract is on drug safety studies, while earlier papers are included to demonstrate research development.

Study design

Apart from one cohort study, all studies used a nested case-control design because this allows for complexity in patterns of exposure and occurrence of risk factors[42]. Nested case-control studies use an underlying cohort structure, in which cases are identified during follow-up and matched controls are randomly selected from all remaining subjects at risk, including potential future cases (incidence density sampling). Cases may serve as controls before their diagnosis date, and controls may be selected for more than one case[43]. This facilitates the extraction of smaller data samples and simpler analysis of time-dependent exposures, and similarities between estimates drawn from nested case-control studies and from their underlying cohorts have been demonstrated[44,45]. In most situations, little statistical power is gained by including more than four or five controls per case[46]. However, for studies with low exposure (less than 15 per cent in controls), particularly when the number of cases is limited, an increase in the control-to-case ratio to ten or more is recommended[47].

Biases

In epidemiological study design using primary care databases, external validity – the generalisability of findings – requires selection and analysis techniques that retain the representativeness of the data, while internal validity – avoidance of skewed results – focuses on biases relating to incomplete recording or systematic differences between

cases and controls[48]. Data acquisition biases applicable to primary care database observational studies fall into two groups: information biases and selection biases.

Information bias and its subcategory, misclassification bias, occur in database studies primarily because some records are not available to researchers when identifying patients. This may arise from confidentiality, the transfer of patients to other health centres, or because some affected patients do not attend a general practice or are not diagnosed or recorded. Two further information bias subcategories – recall and reporting – are not generally an issue because of prospective recording. However, these may affect historical records, such as family histories, because subjects may recall family histories more accurately when diagnosed[49].

Other time-related information biases occur when differential time spans for cases and non-cases in cohort design cause exposure misclassification. Immortal time bias occurs in drug studies when exposed patients have different lengths of non-exposed time (immortal time) prior to exposure, leading to overestimation of associations between exposure and outcome[50]. Using a time-dependent method or nested case-control design addresses this problem by taking into account differential timings of exposure[51].

Immeasurable time bias occurs when exposure records are unavailable, such as when drugs have been administered in hospital for which patients have no associated exposure records, suggesting an incorrect non-exposed status. This reduces the exposure time period for such patients, lowering overall rates or odds ratios[52]. This bias cannot be designed out of such database studies but needs to be discussed as a limitation, with likely effects noted.

Selection bias occurs when the findings are derived partially or entirely from differences between populations of cases and controls. When using primary care

databases, freedom from subcategories – like non-response and survival biases – has been highlighted under database advantages. Selection bias can be avoided by using nested case-control design, in which all participants are drawn from the same cohort, controls are randomly selected and matched to cases, and inclusion and exclusion criteria are applied to all cases and controls.

Other model-related biases may arise during analysis. In randomised control trials, patients are randomly allocated to arms, so patients in all arms have comparable known and unknown risk factors; however, in observational studies, exposed and non-exposed participants may have unequal distributions of risk factors for an outcome. Risk factors associated with, but not a consequence of, outcome and exposure should be considered as confounders in the analysis to avoid confounding bias. Confounding by indication, common in drug safety studies, may occur when a drug is prescribed to treat a medical condition or clinical characteristic associated with an increased risk of an outcome. If not adjusted for, this may lead to overestimation of the association between exposure and outcome[53].

Finally, in observational studies, measurement errors, unknown risk factors, and known risk factors for which there are no data may all cause residual, or uncontrolled, confounding, although the resulting effects will depend on the prevalence of each factor and its scale of association with outcome and exposure[53]. Where they are identifiable and therefore susceptible to discussion, these must be addressed on a study-by-study basis.

Coding of information

Incorrect selection of medical codes for outcomes and confounders may be a source of information bias[54]. So choosing medical codes requires extensive medical expertise and

may involve construction of an algorithm with relevant available data or even validation of the approach using other sources of information or direct interviewing of doctors. Using a code list developed for extraction from one database to identify cases in another may lead to missing cases and a subsequent shifting of any associations with the outcome towards unity. Therefore, if the lists are based on used rather than all codes, a mapping between Read and medication reference tables should be developed to identify the equivalent codes in each database[33].

Combining a primary care database with linked data may add cases which have been under-recorded in general practice. Defining cases and selecting appropriate information is therefore highly complex, requiring general practice expertise and relevant medical knowledge within the research team to assess which information is needed and most reliably recorded in general practices.

Exposure

Researching evidence for causality and associations between exposure and outcome requires designs that incorporate the principles of causality[55]. These include precedence of exposure to outcome, evidence of reversibility of exposure, and evidence of any dose–/duration–response effect.

Exposure is complex and must be summarised appropriately according to the exposure–disease relationship. For some outcomes, toxic instances inhibit recovery of a damaged cell or do not clear and have a cumulative effect over time, while for others, the effect is instantaneous but may also depend on past exposures[56]. Measures of exposure should be developed according to a particular hypothesis, and issues of latency and varying potency, for example for cancer outcomes, should be investigated. For such outcomes, particularly without a strong biological hypothesis, an exposure weighting method may

be used to assess the cumulative effect of drugs at different times[57]. The model includes each unit of exposure with an indicator of timing. This method assumes that each unit continues to damage the tissues and that there is no clearance or repair.

In defining exposure, another exposure-associated bias – protopathic – may be introduced by including prescriptions close to the outcome date for slowly developing conditions such as cancer[58]. Patients may have started or stopped medication only because of changes to their health related to cancer development, thus introducing reverse causality bias. This issue must be considered, and all such prescriptions during an appropriate time period before diagnosis should be removed.

Confounding factors

Confounding factors are an inevitable part of the exposure–outcome model, designed to compensate for non-random allocation to exposure and to reduce residual and indication biases. In primary care database studies, confounders normally include age, sex, ethnicity, lifestyle information (smoking status, alcohol consumption, social deprivation, body mass index), relevant co-morbidities, family history and use of other medication.

Incorrect adjustment for continuous confounders such as body mass index and exact Townsend deprivation scores (available from QResearch® only) will lead to residual confounding. Categorisation (and particularly dichotomisation) should be avoided, as should the assumption of linear relationships between confounders and outcome. Instead, fractional polynomials should be used[59].

To minimise possible protopathic bias caused by diagnoses of co-morbidities due to general ill health or onset of the outcome of interest, comorbidities should be considered only if recorded well before the outcome (at least one year before the cancer diagnosis).

Where family history of an outcome is available, it may be included in analyses, but only if recorded at least six months before the outcome to minimise possible recall bias[60].

Because some potentially important confounder information, such as relevant laboratory test results, stage of cancer, physical activity and diet, is not necessarily consistently recorded in primary care databases, these may not be accounted for in analyses and their likely effects should be included in discussions of results. The extent of the potential effect of biases due to unknown confounders remains a methodological challenge, but this can be assessed, where possible, using sensitivity analyses[61].

Missing values

A multivariate complete-case analysis removes observations with missing values, but this reduces statistical power and may introduce bias. The risk of bias depends on the proportion of missing data and on the reasons why data are missing: completely at random (MCAR), at random (MAR) and not at random (MNAR)[62].

When data are MAR, multiple imputation by chained equations can be used to include all observations in the analysis[63,64]. This process has three steps: generating multiple imputed datasets, analysing these sets, and combining estimates from the sets[65]. First, missing values are replaced by values randomly drawn from their predictive distribution to reflect their uncertainty, each variable being regressed on all other variables to predict its observed values, using logistic, multinomial logistic or linear regression depending on the distribution. The resulting models are used to impute values for observations with missing values, repeating the procedure to create multiple datasets. Next, estimates and variance-covariance matrixes are calculated for each imputed dataset using standard modelling techniques for complete data. Finally, the estimates

are averaged across the datasets, and standard errors estimated using within- and between-imputation variability by applying Rubin's rules[63].

The multiple imputation model needs careful design. Omitting the outcome variable may result in false weakening of the relationship between the outcome and an imputed variable, and may introduce a bias[62]. Included continuous variables are assumed to be normally distributed, so transformations should be applied for highly skewed variables. Considering a wide range of auxiliary variables highly correlated with a variable from which data are missing may reduce the effect of missing information. The number of imputations needed depends on the proportion of missing information. Currently, 20 imputations are recommended for 30 per cent or less of missing information, as is characteristic of lifestyle and body mass index information in the present databases. However, some leeway is possible because the power fall-off with 10 imputations is reportedly only three per cent when compared with 100 imputations[66].

There is no clear way of differentiating between MAR and MNAR, so some form of sensitivity analysis is advisable to assess the extent of bias. In the complete-case approach, if the results of including multiple-imputed data differ from those excluding them, these are reported and the reasons discussed[62]. Another approach – inverse probability weighting complete-case analysis – involves the development of a model for the probability that data are missing. The inverse of the fitted probabilities are then used as weights in the complete-case analysis[67].

Power and multi-database use

Estimating associations between rare adverse effects and relatively new drugs usually requires all available records from large databases, but the number of exposed patients developing a condition of interest is not always sufficient to investigate dose/duration

relationships or drug subtypes. However, accumulating data from more than one database has the potential to overcome this, while possibly also enhancing generalisability, because database data profiles, for example in geographical coverage or differences in recording of information, may be complementary.

To minimise the differences in findings from different databases, the data must be extracted and analysed using the same protocol with comprehensive code lists to cover all included databases. This facilitates application of the fixed-effect model and inverse variance weights method to combine the results. Sensitivity analysis using the random-effect model should be used to detect any heterogeneity.

Designing two identical studies on different data sources with national coverage and then triangulating between the results is a novel approach to research using such databases. It allows a degree of variability between the data samples, but uses an identical study design and code lists that are as similar as possible for data extraction. In addition to achieving more exact and robust estimates, this approach aims also to increase external validity, since the included data may have been collected in slightly different environments and from patients in socially and geographically different areas.

Results

The submitted papers demonstrate the benefits of using contemporary data from sources continually updated over a long period, and applying known and emerging approaches to compensate for lack of data, missing data or biases. Previous studies on the topics covered have been based on older datasets that are more prone to limitations in recording standards, and may reflect less advanced diagnostic knowledge or techniques. Currently available methodologies may also not have been known, fully developed or applied, making the studies more prone to biases. The work submitted exemplifies a move toward research based on the best available data using current techniques to create a standardised, high-quality design. The results demonstrate the potential utility of primary care databases as data sources to deliver generalisable findings from studies of sufficient scale to match the power of traditional attempts, using meta-analytic techniques to improve risk estimates. The three most recent papers also demonstrate the practical benefits achievable through the application of triangulation between the results of identical studies based on QResearch® and CPRD.

Key clinical epidemiological findings

Using all available records, the studies are the largest of their type. Prolonged statin use is found to be associated with increased risks of colorectal, bladder and lung cancers and reduced risks of blood cancers. Prolonged use of COX2 inhibitors shows associations with increased risks of breast and blood cancers and a reduced risk of colorectal cancer. However, bisphosphonate use has no association with increased or reduced risks for any of the common cancers.

The pneumonia studies identify new risk factors for pneumonia, such as stroke and cancer, and show a reduced risk of pneumonia associated with current statin use. The

mental health studies demonstrate increased risks of colon and breast cancer associated with diagnoses of schizophrenia and bipolar disorder, and increased mortality in patients with diabetes suffering from severe mental illnesses.

Considering the large numbers of women using oral contraceptives worldwide, possibly the most important findings of these research studies to date are the associations between use and risks of VTE for various oral contraceptive drug types and exposures. There has been frequent publicity about concerns relating to more recent formulations introduced to reduce other side-effects. While the submitted study conclusively demonstrates the relatively low VTE risks of all oral contraceptives when compared with pregnancy, the findings do confirm a higher incidence of VTE events in newer generation drugs compared with older formulations.

All the studies provide new epidemiological evidence, adding knowledge to the findings of earlier research, and confirming, negating or arbitrating between conflicting results. In all cases, the studies provide greater certainty, and often add detail to earlier research. Many results provide reassurance to doctors and patients regarding the safety of preventive prescribing over long periods, while a few reported risk associations should inform prescribing decisions for patients in particular risk categories. All the findings are potentially useful for people developing treatment policies and for those prioritising pharmacoepidemiological or drug development research.

Study design

Apart from the single cohort study (Paper 10), which was based on patients with diabetes, all other studies were based on the general population and so used nested case-control design. Each case was matched with up to five controls by year of birth and gender. The controls were alive and registered with the same practice at the time of

their matched case diagnosis or index date. In the drug studies, a range of medications was available to treat investigated symptoms or conditions, so practice-specific preferences for a subset might develop. Matching by practice was used to reduce residual confounding by increasing the likely similarity of prescribing for cases and their controls. Similarly, nested case-control matching by calendar year minimised residual confounding from changing rates of exposure to a drug of interest over the study period.

Matching by practice and calendar year restricted the number of controls matched by age and sex. In the contraceptive and VTE study (Paper 8), only about 30 per cent of cases had five controls. Although exposure to some drugs was relatively low, increasing the number of controls would not be useful. In theory, ten controls would increase the power from 90 to 94 per cent (Paper 7). In reality, however, only a small proportion of cases would have the larger number of controls, so the power gain would be much smaller.

In the power calculations for the bisphosphonate study (Paper 4), 4.2 per cent of patients were assumed to be exposed. The numbers of available cases for the cancer of most concern, oesophageal, were 5,364 in QResearch® and 5,132 in CPRD. Changing the number of matched controls from five to ten would have increased the power to detect an odds ratio of 1.3 from 87 to 90 per cent in QResearch® and from 85 to 89 per cent in CPRD. However, around 30 per cent of cases had fewer than five matched controls already, with no further appropriate controls available for matching, so requiring more controls would have resulted in a much smaller power increase.

Study population

Apart from the cohort study of patients with diabetes, all studies were based on the general population registered with the contributing practices. This approach was challenged during the peer-review process for the study of contraceptive drugs (Paper 8) by a suggestion that only healthy women (without chronic and acute conditions) should be included in the study. The reviewer argued that acute events (trauma, operation, infection) are proximate causes rather than confounders. However, because a woman would be likely to stop taking contraceptive drugs in the case of an acute event, and particularly before a scheduled surgery, it was decided that these variables were more properly treated as confounders. The same reviewer further suggested that "when one wants to evaluate the independent effect of a drug on an outcome it is best to study the effect in a generally healthy population". This approach has been used in previous studies, but without any consistent definition of proximal causes of VTE. Excluding these observations would also lead to non-generalisable results and overestimation of risks associated with contraceptive use.

Outcomes and exclusions

All the submitted studies were affected by a lack of adjudicated outcomes and, apart from the latest study (Papers 7 and 8), had access to GP records only, without the availability of linked data on hospital and post mortem diagnoses. In the cancer studies (Papers 1 to 6), the incidence rates based only on GP records may have been slightly lower[38] because some cases had been registered only at time of death and not recorded by the general practice. In the pneumonia studies (Papers 9 and 10), patients used as controls might also have died of pneumonia with no record placed in their practice records. However, the effect of misclassification bias arising from such errors in case identification will be small because cancer and pneumonia are rare in the general

population and even rarer in the residual population once known cases have been removed. Over-recording of outcomes was potentially important in the pneumonia studies (Papers 9 and 10), because outcome diagnosis was based on GP records of clinical symptoms, without requiring diagnostic support from X-ray examinations or microbiological confirmation. However, the pneumonia incidence rates in these studies were in line with commonly published ranges, and a sensitivity analysis using only cases verified by hospitalisation suggested no sizeable contribution to information bias.

For the cancer studies, selection of cases was also based on the first recorded diagnosis, while the exact site may have been determined only later. This lack of distinction between types of cancer may have led to under-recording of some outcomes and over-recording of others. Similarly, information about stage of cancer was not consistently recorded and could not be used. Furthermore, the studies did not have access to information about cancer screening tests (mammography, prostate-specific antigen or colonoscopy), so some patients may have been wrongly included, having already had a cancer.

Under- and over-recording of outcomes were also relevant in the VTE study (Papers 7 and 8), but under-recording from direct admissions to hospital was resolved using links to HES and ONS mortality data. For over-recording, some earlier VTE studies based on CPRD data had used pseudo-verification of the diagnosis by requiring consequent prescription of anti-coagulation therapy. This would be appropriate in health systems where information for all prescriptions is available. However, in the UK, anticoagulant prescriptions issued by a GP may simply reflect the initial concern of a doctor for patients with a known possible reason, where prescription may pre-date confirmatory tests in an anticoagulant clinic. Moreover, consequent prescriptions in secondary care do not appear in GP records.

Exposure

All analyses considered no use of the drug of interest as the reference category. Exposure was based on recorded prescriptions and, in most studies, exposure to a drug of interest was defined as at least one prescription before the index date. To assess the duration–response effect, cumulative exposure was estimated by extracting every prescription, grouping all prescriptions with small inter-prescription gaps, taking grouped prescription course times as from the beginning of the first to end of the last, and calculating cumulative duration as the sum of all overall course times for each patient. The minimum length of gap between courses was drug- and outcome-dependent: 30 days for oral contraceptives looking for short-term VTE outcome; 60 days for exposure to statins or COX2 inhibitors; and 90 days for exposure to bisphosphonates, which accumulate in bones and gradually release.

Duration was categorised for reporting and comparability purposes, but trend tests were run on actual months of exposure. Reversibility of exposure, which is particularly important for acute conditions such as VTE and pneumonia, was assessed by analysing the gap between the end of the last prescription and the index date. The effect of past exposures was assessed only in the last cancer study (Papers 5 and 6). Since the available data did not allow greater analytic detail, the interaction of timing and terms of treatment were categorised as: short-term (less than 12 months) recent use (within two years of the index date); short-term remote use (at least two years before the outcome); long-term (more than 12 months) recent use; and long-term remote use.

Prescription records do not guarantee actual use so may introduce information bias due to misclassification of exposure, particularly if patients with only one prescription are included. Adherence to statins is not high, with an estimated drop-out rate reported in one study of about 30 per cent six months after inception[68]. To increase the likelihood

of patients being actual users in statin studies (Papers 1, 2 and 12), exposure was defined as at least two prescriptions in the observation period. However, bisphosphonates have a long-term effect, and even one prescription may be important, so this study (Papers 5 and 6), defined exposure as at least one prescription for the main analysis, and included a sensitivity analysis redefining exposure as at least two prescriptions to assess possible misclassification bias. Similar results were obtained, confirming the expectation of no systematic difference in adherence between cases and controls. Another misclassification of exposure arises from drugs obtained from sources other than general practices. For the oral contraceptives study, these included both contraceptive clinics, used by an estimated 6.3 per cent of women under 25 and 1.2 per cent of older women[69], and pharmacies. However, additional stratified analysis for these age groups demonstrated associations similar to the main analysis.

Confounding factors

In the submitted studies, some confounders were established *a priori*, while others were included in the analysis as established risk factors for outcomes, thus potentially affecting doctors' prescribing decisions. Specifically, in the VTE study, a list of relevant chronic and acute conditions was derived from NHS guidelines, and where there were records before the index date – for acute conditions, only those between six months and one month before – these were included (Paper 8). In all cancer studies, comorbidities were considered only if recorded at least one year before the index date to minimise protopathic bias. The bisphosphonate study (Papers 5 and 6), included gastrointestinal disorders only if diagnosed before the first use of bisphosphonate (12 months before the index date for non-users) to avoid confounding bias and misleading associations between bisphosphonates and oesophageal cancer.

The quality and availability of lifestyle variables (smoking status, BMI, alcohol consumption and social deprivation) increased over time. To minimise residual confounding, the most detailed information available at the time was used. Categories for smoking, alcohol consumption and ethnicity were collapsed, but only to achieve sufficient observations per category. In earlier studies, the BMI of many patients was recorded only as a category (normal, overweight or obese), but later studies benefited from better recording and used exact BMI values.

Ethnicity information was poorly recorded in the past. However, the quality is continually improving and it was used in the later Papers 5, 6 and 8. It was still only incompletely recorded, so the 'not recorded' category together with 'white' was used as the reference category, rather than a specific ethnic group. Adjusting for ethnicity goes some way to addressing possible differential cancer[70] and VTE risks[71], as well as effects from different lifestyles and prescribing patterns[72].

Analysis

Models

Multivariate models appropriate to the study design were used. For the cohort study (Paper 10), which investigated associations between various factors and mortality, the Cox model was selected, but all others with nested case-control design used conditional logistic regression. Inclusion of a continuous variable for BMI, assuming a linear relationship with the outcome, might have contributed to residual confounding. This could be improved by using fractional polynomials. All analyses started from univariate models assessing the association between the outcome and the exposure or co-morbidity of interest, with effects from adding confounding factors carefully controlled.

The data extracted from the databases were very similar and the only adjustment required to any model occurred in the bisphosphonates study. The CPRD analysis included one extra category of ex-users for alcohol consumption, whereas QResearch® data had insufficient patients for this category (Papers 5 and 6).

Missing values

Owing to numerous missing values for BMI (about 25 per cent), smoking (11-15 per cent) and alcohol consumption (23 per cent), a category of 'not recorded' was used in the earlier studies (Papers 1, 9 to 12), thus adding to residual confounding bias. In the later studies (Papers 2 to 8), on the assumption that data were MAR, multiple imputation was used. In general, data for BMI, smoking and alcohol consumption are not missing completely at random, but including information such as age, sex, co-morbidities, various medications and outcomes in the imputation models does make the MAR assumption plausible. All studies used sensitivity analyses of complete data to assess this assumption of MAR and demonstrated similar results.

Residual confounding

Several potential uncertainties in the submitted studies arose from lack of or deficiencies in the data, requiring assumptions to be made. To counteract these, sensitivity and additional analyses were run. Some were run on sub-samples to confirm the results obtained from the main analyses. For example, in the pneumonia studies, analyses were run on sub-samples of patients with the most likely verified outcome. All studies were checked for accuracy of data by analysis of a sub-sample with more years of records. Additional analyses were run for specific sub-samples to estimate the effect of variations in the design and to facilitate comparisons with other studies. However, none of the studies was assessed for the potential effect of unobserved confounding, and sensitivity analyses to assess this would have been valuable.

Power and multi-database use

The last two projects used both QResearch® and CPRD (Papers 5 to 8). Separate studies were carried out for each database using the same protocol. Sample sizes were very similar, so differences in observed associations were likely to be caused only by sampling variation. A fixed effect-model and an inverse variance weights method was therefore used to combine the results, together with a sensitivity analysis using the random-effect model to allow for any heterogeneity.

In the first project, on cancer and bisphosphonates, for most cancers the results from QResearch® and CPRD were very similar, with overlapping confidence intervals (Papers 5 and 6). For the second project, on VTE and contraceptives, the main challenge was the low use of newer contraceptives (Paper 8). Most of the results were very close, so applying a fixed-effect model to combine the results was justifiable. For one drug, Norgestimate, the confidence intervals from the databases did not overlap (1.96, 1.56 to 2.46 for CPRD and 3.15, 2.56 to 3.89 for QResearch®), indicating significant heterogeneity (I^2=89%, P=0.003) in these findings. Because there were no clear reasons for differences in exposures in practices from QResearch® and CPRD and the direction of the effect was the same in both databases, the results from the fixed-effect model were reported as the main finding (2.53, 2.17 to 2.96). However, a random-effect model was run as a sensitivity analysis to allow for the found heterogeneity. In both studies, triangulation of the results delivered more precise estimates, with on average 30 per cent narrower confidence intervals based on a larger and more generalisable population.

Discussion

Taken together, the submitted papers and this extended abstract represent a contribution to epidemiological knowledge, primarily in the area of drug safety. They demonstrate the importance of using emerging and developing methodological techniques to address data weaknesses, thereby maximising the accuracy and reliability of results. The approach of combining findings from identical separate analyses on two databases is also a novel extension to methodological practice in this area of research.

Strengths and weaknesses of the presented research

Achieving the power necessary to produce accurate estimates is a common problem when researching rare medical outcomes. This issue assumes even greater importance when there is a requirement for detailed research, such as distinguishing between drug types, dosage regimes, or a combination of the two. The generalisability of results, in terms of whether they are representative of real outcomes in the general population, is another issue important to epidemiological research. Both are requirements that only data sources like the large primary care databases are able to satisfy.

The routine, prospective collection of QResearch® and CPRD data freed the submitted studies of selection, recall and responder bias. Including HES- and ONS-linked data in the most recent paper added more cases and further reduced misclassification bias. The nested case-control design studies included all available cases, overcoming restrictions on the amount of data extracted. To allow for differences between exposed and non-exposed patients, all available confounders were included in the analyses. In general, the submitted studies included more confounders, and more systematically applied, than previous published research. Exposure was also defined to facilitate investigation of variation in potency and latency, and to provide more detailed estimates for clinicians

and future researchers. Using multiple imputation to account for uncertainty associated with missing data further improved the validity of the findings of the later studies.

The data used have limitations. Historically, limited allowance has been made for specific research needs because the data have simply been routinely collected as events occur (one exception to this is where, in QResearch®, EMIS integrated the postcode deprivation table into its system to allow extraction of deprivation scores without the postcode to protect patient confidentiality). Some useful information may not have been recorded, or may have been recorded in ways that cause difficulties of interpretation for researchers, occasionally contributing to residual confounding. There are also inconsistencies and selectivity in recording, with not all records being updated in a timely fashion, contributing to information bias. As noted before, available data are based on prescriptions issued rather than actual use, introducing misclassification bias. All these issues affect estimates in an unquantifiable way, even when techniques to adjust for these or to estimate their effect are applied.

Multi-database use

Traditionally, meta-analyses have been used to increase power by combining the results of studies with similar research questions. A key weakness of this approach is that the studies included have often been conducted in different environments, and have used different designs, sample selection, study periods, exposure definitions and sets of confounders. This inevitably introduces additional uncertainty, expressed in wide confidence intervals, and the results are unlikely to be truly representative of the general population.

In the last of the papers submitted, the inherent benefits of large-scale, well-designed studies using a single primary care database was extended by applying a meta-analytic

technique to triangulate between the results from QResearch® and CPRD. This is in effect a meta-analysis, but between two identical studies based on two large datasets with a high degree of commonality in terms of data collection, curation and purpose, and avoiding the problems of study disparities found in traditional meta-analysis. The methodological contribution lies in the successful use of a meta-analytic approach to extend the power achievable by even the largest current primary care database. Although in practice the combined power of QResearch® and CPRD was sufficient for the requirements of this research, the method could, in principle, be used to include the results of many similar data sources.

Triangulation of this kind usefully combines results and delivers better estimates but must be applied with care. It does not resolve issues arising from underlying data quality: if the separate studies are prone to a specific shift in estimations from a number of shared biases, this will result in more precise, but still biased, combined estimates. This possibility places greater responsibility on researchers designing such studies. Other limitations may arise from differences in data recording between the data sources used, such as exact values versus categorisation, or the amount of detail given. Although in most cases the effect is likely to be small, it may require slight differences in modelling, which needs to be considered and fully explained.

In essence, the triangulation technique used was a two-step meta-analytic approach, in which individual participant data were analysed separately and then combined[73]. This would be the only approach possible if data from some datasets had to be analysed in-house and were not available to all research groups in a geographically dispersed team. However, combining data aggregated in this way does not allow modelling of exposure risks across all observations. This would restrict the assessment of risks for rarer types of drugs and the magnitudes of modifying effects of rarer patient characteristics.

Further development of the triangulation approach could derive from application of a one-step meta-analytic approach. This would involve the pooling of individual data while preserving patient clustering within the databases, and might provide not only even greater power but also improvements in the quality of the findings. The two approaches have been found to produce similar results, particularly when estimating associations between a single treatment and outcome[73,74]. Pooling individual data would allow analyses to be run across all observations, delivering estimates for rare types of exposure and adjustments for rare confounders.

Such an approach requires standardisation of the data so that the combined datasets have the same variables of similar types. This would make it applicable to designs such as cohort studies, and to analyses adjusted for baseline confounders and time-dependent exposures. It might also facilitate the development of prognostic models. Riley highlights the complexity and effort required when this approach is applied across a variety of studies independently undertaken by many separate teams[74]. Applying this technique on a few bespoke, effectively identical studies using very large datasets with a high degree of similarity should, however, significantly improve efficiency and effectiveness.

Methodological issues

Because observational studies may be subject to various biases, it is important to ensure transparency in the reported findings. General guidelines for observational studies are listed in the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement and are endorsed by medical journals[75]. The STROBE checklist includes full descriptions of the data and analysis, results and limitations of a study, and further extensions of the statement have been developed for specific types of observational studies.

To address specific issues arising from routinely collected data, an international group of scientists has introduced a further guideline for such studies: the Reporting of Studies Conducted using Observational Routinely-Collected Health Data (RECORD) statement[54]. In addition to STROBE items, this deals with misclassification biases and requires reporting of the complete list of codes and algorithms for population selection and for outcomes, exposures, confounders and effect modifiers. With respect to data quality, the statement prompts researchers to describe any filtering, data availability and linkage. The limitations section must also include discussion of unmeasured confounding, missing data, changing eligibility over time and misclassification bias, all of which are associated with routinely collected data. Methodologies for overcoming these limitations remain analytical challenges requiring further research.

In summary, the following are key issues for researchers:

- When designing observational studies, it is essential to consider all possible biases. In addition to the practice records, linked data should be used to identify as many patients with outcomes as are available and to exclude patients with outcomes prior to the study period. When using a case-control design, it is also important to have sufficient controls to ensure accurate estimates.
- As much information detail as possible should be retained for confounding factors, for example by using fractional polynomials to describe complex continuous variables. All possible issues regarding unmeasured confounders should be addressed by designing sensitivity or additional analyses. The effect of unobserved confounders should also be estimated.
- To handle missing values, possible reasons for missing data should be considered. Potential MAR data should be multiply imputed and the imputation model should

include the outcome and as many relevant variables as possible. A sensitivity analysis should also be run to assess the plausibility of the MAR assumption.

- Analyses should be run on imputed datasets separately for each database. The results obtained from all analyses should be combined using the inverse variance weights method and, after testing the assumption of no heterogeneity, applying the fixed-effect model.
- Where suitable databases are available, using more than one database may be beneficial, particularly for rare conditions or complex exposures. Appropriate measures should be taken to test for heterogeneity assumptions. When multiple databases are used, the results of the separate analyses should be presented along with the combined findings.
- Where more than one database is being used and it is possible to use the individual-level data together, rather than running separate analyses on the two datasets and combining the result, it might be beneficial to pool the data and run the combined analysis across all observations, clustered by database.

Primary care databases in the UK

The overview of the current environment shows the extent to which primary care data collection has developed in the UK, as well as the growing exploitation of these data for care and research purposes. The background to these developments has been complex, and the drivers for them various.

Regional and national primary care databases are defined by their coverage. They collect as complete a picture as possible of general healthcare in their area of interest, using data encoded as part of normal clinical practice, with minimal intervention apart from feedback to contributing practices. The value of these resources lies in their ability to reflect real-life practice and produce results generalisable across the population

defined by their coverage. The larger they are, the more utility they will have for studies of rare diseases, outcomes and drug exposures, or to support investigations into, for example, more detailed exposure.

Smaller, more local databases are usually associated with more focused aims, particular conditions or treatments. Their smaller size facilitates a greater level of intervention in the collection and curation of data to maximise data completeness and accuracy in areas of specific interest. However, their use is limited to in-depth studies of commoner conditions, where more detailed data, sometimes augmented by data directly input by patients, are optimal.

Between database types, there does not appear to be a ‘better’ or ‘best’ approach. Small specialised databases will continue to be most appropriate for specific research problems. Through incremental growth and continual improvements to the range and quality of data stored, regional and national databases can and will provide sound bases for studies involving rarer conditions and rarer treatments or where outcomes for a population are required.

Developments in database content

The existing Read 2 coding used in QResearch® and CPRD has a number of internal problems. Parts of the vocabulary are full and some terms are inaccurate. Most importantly, however, the system is not used in hospitals. Electronic patient data are recorded and shared across all care settings, so a consistent coding system would clearly be beneficial. It would decrease the effort required to enter information, eliminate the need for recoding, reduce risks of error when linking or pooling data, and make coding a skill transferable across the care environment.

Fortunately, a new system to replace all versions of Read code usage and vocabularies is planned for adoption by the entire health system by 2020. Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT®) is an internationally developed hierarchical coding system, which will reduce the need to augment coding with free text[76]. The system will integrate the NHS Dictionary of Medicines and Devices, and for the first time provide a link between clinical information and prescribing data. The UK Terminology Centre already provides presentations and webinars (http://systems.hscic.gov.uk/data/uktc/training) and other information on SNOMED CT. The introduction of SNOMED CT, the provision of appropriate system interfaces for different medical settings, and consistent staff training could bring about major improvements in data completeness and accuracy.

This might also facilitate other developments. Some information about events occurring outside of general practices, such as prescriptions issued, is currently unavailable to researchers, and it is also impossible to know whether a patient actually uses any issued prescription. Links to prescribing information in hospitals and clinics and to drug issuing information from pharmacies would be valuable steps toward the completeness of primary care data and would facilitate more accurate estimates of drug use, safety and effectiveness.

The current lack of, or failure to regularly update, some personal and health-related patient characteristics important to research also impacts on the utility of large primary care databases that rely on the quality of routine recording in general practice and other primary care settings. Common information gaps important for research use include changing or fluctuating measures such as smoking status, weight and blood pressure.

Currently, personal and lifestyle information is most likely to be recorded only when a patient joins a practice or when a doctor is assessing risks associated with, for example,

smoking or BMI. This weakness could be largely eliminated by regular annual or more frequent visits for all registered patients. Currently, cost and other pressures might make this unwelcome but, as part of a wider preventative medicine initiative, it could prove very valuable.

Improvements in the completeness and quality of primary care data might also be effected through patient online access to practice medical records (http://www.hscic.gov.uk/pomi). The NHS has introduced an indicator to assess the availability of booking, ordering of repeat prescriptions and access to medical records online. It might be useful to allow patients also to check and challenge records, or to enter structured information about other personal characteristics, such as education, occupation, physical and emotional state (stress or lack of sleep), physical activity, diet, compliance with currently prescribed medicine, and use of unprescribed medications or remedies. Information about actual use of prescribed or other drugs, if ever recorded, is also not coherently organised, and allowing patients to enter, review and update such information might over time help to improve the completeness of primary care data available for research.

Database development and strategy

In 2013, the national care.data programme was initiated by the NHS and the Health and Social Care Information Centre to create a central database linking anonymised health and social care records (www.england.nhs.uk/ourwork/tsd/care-data/). This project caused extensive debate about data security and the need for further development of data security standards[77]. In July 2016, the NHS closed the project. The problems could have been anticipated, especially in a country where the privacy of healthcare records has long been a cornerstone of many individuals' value systems. The Internet has created an environment in which state- and private-sector intrusiveness into personal

data is a matter of political debate, fuelled by regular, well-publicised ‘hacking’ events involving supposedly secure government and private-sector systems. Any national system for the systematic universal collection of personal data is therefore likely to fail because of political and privacy issues.

Given the size and complexity of the environment and the growing competition for resources, overall improvements for researchers requiring large-scale data resources will, therefore, probably best be delivered by initiatives to enrich the existing different resources, especially with regard to the range and quality of data stored, to make them more complementary. This thesis demonstrates the feasibility of using triangulation to exploit multiple resources in the pursuit of a single study topic. Developments in coding consistency and recording practice, like those outlined above, might be used to increase the inter-operability of different data sources. An environment in which research teams could use multiple similarly-constructed databases to achieve high levels of coverage from different care locations across the whole or any part of the UK would be a major advance, facilitating very large studies with the power to investigate problems where data sparsity or consistency currently remains an issue.

Such multi-database studies might mitigate unevenness of coverage, between England and other parts of the United Kingdom, and possibly also across more remote areas. The current national nature of hospital records is a particular problem, resulting in the absence of linkages to any equivalents for English HES data in Scotland and Wales. Initiatives to address this would clearly be advantageous. However, this may not be politically or economically feasible at the current time.

Conclusion

Observational studies using data from routinely-collected large-scale primary care information sources can deliver useful contributions to clinical epidemiological areas such as drug safety and disease risk factors. They provide the benefit of enabling real-life outcomes of normal clinical practice in the general population to be assessed, and may include rare or slowly-developing conditions, or analysis of treatments at the level of drug type, different exposure terms or drug regimes.

The submitted work and extended abstract also show the progress possible when emerging/developing methodologies are applied to address problems associated with such data sources. The abstract uses the experience gained to provide recommendations to researchers wishing to use such resources, and to make observations about improvements to their content and possible developments in primary care data resources.

The benefits of a novel use of a meta-analytic technique to combine results from two very similar large databases to increase statistical power and the quality of findings have been demonstrated. Triangulation between the results of separate identical studies on similar data sources has significant advantages over more traditional meta-analyses, in which the incorporated studies and datasets are often very disparate. Possible further development of the approach has also been mooted.

References

1. Abrahamsen B, Pazianas M, Eiken P, Russell RGG, Eastell R. Esophageal and gastric cancer incidence and mortality in alendronate users. *J Bone Miner Res.* 2011;27:679-686.
2. Lidegaard Ø, Nielsen LH, Skovlund CW, Skjeldestad FE, Løkkegaard E. Risk of venous thromboembolism from use of oral contraceptives containing different progestogens and oestrogen doses: Danish cohort study, 2001-9. *BMJ.* 2011;343:d6423.
3. Jackevicius CA, Li P, Tu JV. Prevalence, predictors, and outcomes of primary nonadherence after acute myocardial infarction. *Circulation.* 2008;117(8):1028-1036.
4. Penning-van Beest FJA, Termorshuizen F, Goettsch WG, Klungel OH, Kastelein JJP, Herings RMC. Adherence to evidence-based statin guidelines reduces the risk of hospitalizations for acute myocardial infarction by 40%: a cohort study. *Eur Heart J.* 2007;28(2):154-159.
5. Abraha I, Montedori A, Stracci F, Rossi M, Romagnoli C. Statin compliance in the Umbrian population. *Eur J Clin Pharmacol.* 2003;59(8-9):659-661.
6. Halava H, Korhonen MJ, Huupponen R, et al. Lifestyle factors as predictors of nonadherence to statin therapy among patients with and without cardiovascular comorbidities. *CMAJ.* 2014;186(12):E449-E456.
7. Lee W-Y, Sun L-M, Lin M-C, et al. A higher dosage of oral alendronate will increase the subsequent cancer risk of osteoporosis patients in Taiwan: a population-based cohort study. *PLoS ONE.* 2012;7(12):e53032.
8. Bird ST, Hartzema AG, Brophy JM, Etminan M, Delaney JAC. Risk of venous thromboembolism in women with polycystic ovary syndrome: a population-based matched cohort analysis. *Can Med Assoc J.* 2013;185(2):E115-E120.
9. Gronich N, Lavi I, Rennert G. Higher risk of venous thrombosis associated with drospirenone-containing oral contraceptives: a population-based cohort study. *Can Med Assoc J.* 2011;183(18):E1319-E1325.
10. Parkin L, Skegg DCG, Wilson M, Herbison GP, Paul C. Oral contraceptives and fatal pulmonary embolism. *Lancet.* 2000;355(9221):2133-2134.
11. Ruigómez A, Brauer R, Rodríguez LAG, et al. Ascertainment of acute liver injury in two European primary care databases. *Eur J Clin Pharmacol.* 2014;70(10):1227-1235.
12. Secretary of State for Health. *The NHS Plan, a Plan for Investment, a Plan for Reform.* London: The Stationery Office; 2000.
13. Griffiths F, Wild A, Harvey J, Fenton E. The productivity of primary care research networks. *Br J Gen Pract.* 2000;50(460):913-915.
14. Barnes H, Pilling D. Stimulating R&D in primary care. *Prim Health Care.* 2003;13(3):38-38.
15. Chisholm J. The Read clinical classification. *BMJ.* 1990;300:1092.

16. Tate AR, Martin AGR, Ali A, Cassell JA. Using free text information to explore how and when GPs code a diagnosis of ovarian cancer: an observational study using primary care records of patients with ovarian cancer. *BMJ Open.* 2011;1(1): e000025.

17. RCGP. *The Good Practice Guidelines for GP Electronic Patient Records*. Department of Health; 2011.

18. Moffat K, McNab D. Improving management of gout in primary care using a customised electronic records template. *BMJ Qual Improv Report.* 2015;4(1): u204832.w2038.

19. Townsend P. Deprivation. *J Soc Policy.* 1987;16:125-146.

20. Rowan S. Trends in cancer incidence by deprivation, England and Wales, 1990-2002. *Health Stat Q*;36:24-35.

21. Hawker J, Olowokure B, Sufi F, Weinberg J, Gill N, Wilson R. Social deprivation and hospital admission for respiratory infection: an ecological study. *Respir Med.* 2003;97(11):1219-1224.

22. Gregory IN. Comparisons between geographies of mortality and deprivation from the 1900s and 2001: spatial analysis of census and mortality statistics. *BMJ.* 2009;339:b3454.

23. Yu D, Peat G, Bedson J, Jordan KP. Annual consultation incidence of osteoarthritis estimated from population-based health care data in England. *Rheumatology.* 2015;54(11):2051-2060.

24. Brusselle G, Price D, Gruffydd-Jones K, et al. The inevitable drift to triple therapy in COPD: an analysis of prescribing pathways in the UK. *Int J Chron Obstruct Pulmon Dis.* 2015;10:2207-2217.

25. Jick SS, Kaye JA, Vasilakis-Scaramozza C, et al. Validity of the General Practice Research Database. *Pharmacotherapy.* 2003;23(5):686-689.

26. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK primary care data resource. *Ther Adv Drug Saf.* 2012;3(2):89-99.

27. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract.* 2010;60:e128-e136.

28. 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(5):443-453.

29. Lawrenson R, Williams T, Farmer R. Clinical information for research: the use of general practice databases. *J Public Health Med.* 1999;21(3):299-304.

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

31. Hippisley-Cox J, Hammersley V, Pringle M, Coupland C, Crown N, Wright L. Methodology for assessing the usefulness of general practice data for research in one research network. *Health Inform J.* 2004;10:91-109.

32. ICO. *Anonymisation: Managing Data Protection Risk Code of Practice Summary*. Wilmslow: Information Commissioner's Office; 2012.

33. Hippisley-Cox J, Coupland C, Brindle P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. *BMJ Open.* 2014;4(8):e005809.

34. Reeves D, Springate DA, Ashcroft DM, et al. Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with ischaemic heart disease: a cohort study with nested case-control analysis. *BMJ Open.* 2014;4(4):e004952.

35. Bhaskaran K, Forbes HJ, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). *BMJ Open.* 2013;3(9):e003389.

36. Booth HP, Prevost AT, Gulliford MC. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. *Pharmacoepidem Dr S.* 2013;22(12):1357-1361.

37. Delaney JAC, Moodie EEM, Suissa S. Validating the effects of drug treatment on blood pressure in the General Practice Research Database. *Pharmacoepidem Dr S.* 2008;17(6):535-545.

38. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol.* 2010;69(1):4-14.

39. Wright E, Schofield PT, Seed P, Molokhia M. Bisphosphonates and risk of upper gastrointestinal cancer: a case control study using the General Practice Research Database (GPRD). *PLoS ONE.* 2012;7(10):e47616.

40. Green J, Czanner G, Reeves G, Watson J, Wise L, Beral V. Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort. *BMJ.* 2010;341:c4444.

41. Walker AJ, Card T, Bates TE, Muir K. Tricyclic antidepressants and the incidence of certain cancers: a study using the GPRD. *Br J Cancer.* 2011;104(1):193-197.

42. Etminan M. Pharmacoepidemiology II: The nested case-control study – a novel approach in pharmacoepidemiologic research. *Pharmacotherapy.* 2004;24(9):1105-1109.

43. Lubin JH, Gail MH. Biased selection of controls for case-control analyses of cohort studies. *Biometrics.* 1984;40(1):63-75.

44. Langholz B, Thomas D. Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison. *Am J Epidemiol.* 1990;131(1):169-176.

45. Wacholder S. Bias in full cohort and nested case-control studies? *Epidemiology.* 2009;20(3):339-340.

46. Rothman K, Greenland S. *Modern Epidemiology,* vol 1., 2nd ed. Boston: Lippincott-Raven; 1998.

47. Hennessy S, Bilker WB, Berlin JA, Strom BL. Factors influencing the optimal control-to-case ratio in matched case-control studies. *Am J Epidemiol.* 1999;149(2):195-197.

48. Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Commun H.* 2004;58(8):635-641.

49. Chang ET, Smedby KE, Hjalgrim H, Glimelius B, Adami H-O. Reliability of self-reported family history of cancer in a large case-control study of lymphoma. *J Natl Cancer I.* 2006;98(1):61-68.

50. Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol.* 2008;167(4):492-499.

51. Suissa S. Novel approaches to pharmacoepidemiology study design and statistical analysis. In: Strom BL, ed. *Pharmacoepidemiology*. London: John Wiley & Sons Ltd; 2005:811-829.

52. Suissa S. Immeasurable time bias in observational studies of drug effects on mortality. *Am J Epidemiol.* 2008;168(3):329-335.

53. Psaty BM, Koepsell TD, Lin D, et al. Assessment and control for confounding by indication in observational studies. *J Am Geriatr Soc.* 1999;47(6):749-754.

54. Benchimol EI, Smeeth L, Guttmann A, et al. The Reporting of Studies Conducted using Observational Routinely-Collected Health Data (RECORD) statement. *PLoS Medicine.* 2015;12(10):e1001885.

55. Hill A. The environment and disease: association or causation? *Proc R Soc Med.* 1965;58(5):295-300.

56. Kriebel D, Checkoway H, Pearce N. Exposure and dose modelling in occupational epidemiology. *Occup Environ Med.* 2007;64(7):492-498.

57. Checkoway H, Pearce N, Kriebel D. Exposure and dose modelling. In: *Research Methods in Occupational Epidemiology*, vol 34. Oxford: Oxford University Press; 2004.

58. Tamim H, Monfared AAT, LeLorier J. Application of lag-time into exposure definitions to control for protopathic bias. *Pharmacoepidem Dr S.* 2007;16(3):250-258.

59. Groenwold RHH, Klungel OH, Altman DG, van der Graaf Y, Hoes AW, Moons KGM. Adjustment for continuous confounders: an example of how to prevent residual confounding. *Can Med Assoc J.* 2013;185(5):401-406.

60. Mai PL, Garceau AO, Graubard BI, et al. Confirmation of family cancer history reported in a population-based survey. *J Natl Cancer Inst.* 2011;103(10):788-797.

61. Groenwold RHH, Hak E, Hoes AW. Quantitative assessment of unobserved confounding is mandatory in nonrandomized intervention studies. *J Clin Epidemiol.* 2009;62(1):22-28.

62. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393.

63. Royston P. Multiple imputation of missing values. *Stata J.* 2004;4(3):227-241.

64. Royston P. Multiple imputation of missing values: update of ice. *Stata J.* 2005;5(4):527-536.

65. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011;30(4):377-399.

66. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science.* 2007;8(3):206-213.

67. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res.* 2013;22(3):278-295.

68. Simons LA, Simons J, McManus P, Dudley J. Discontinuation rates for use of statins are high. *BMJ.* 2000;321(7268):1084.

69. HSCIC. *NHS Contraceptive Services: England, 2004/05*. London: Health and Social Care Information Centre; 2005.

70. Ferlay J, Shin H-R, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer.* 2010;127(12):2893-2917.

71. White RH, Zhou H, Romano PS. Incidence of idiopathic deep venous thrombosis and secondary thromboembolism among ethnic groups in California. *Ann Intern Med.* 1998;128(9):737-740.

72. French R, Mercer C, Johnson A, Fenton K, Erens B, Wellings K. Use of contraceptive services in Britain: findings from the second National Survey of Sexual Attitudes and Lifestyles (Natsal-2). *J Fam Plann Reprod Health Care.* 2009;35(1):9-14.

73. Mathew T, Nordström K. Comparison of one-step and two-step meta-analysis models using individual patient data. *Biometrical J.* 2010;52(2):271-287.

74. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ.* 2010;340.

75. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *PLoS Med.* 2007;4(10):e296.

76. UK Terminology Centre. *SNOMED CT® A User Guide for General Practice*. London: NHS; 2012: www.nhscfh.nhs.uk/uktc.

77. Caldicott F. *Review of Data Security, Consent and Opt-Outs.* London: National Data Guardian for Health and Care; 2016.

Published works

Paper 1

Vinogradova Y, Hippisley-Cox J, Coupland C, Logan R. Risk of colorectal cancer in patients prescribed statins, nonsteroidal anti-inflammatory drugs, and cyclooxygenase-2 Inhibitors: nested case-control study. *Gastroenterology*. 2007;133:393–402.

CLINICAL–ALIMENTARY TRACT

Risk of Colorectal Cancer in Patients Prescribed Statins, Nonsteroidal Anti-Inflammatory Drugs, and Cyclooxygenase-2 Inhibitors: Nested Case-Control Study

YANA VINOGRADOVA,* JULIA HIPPISLEY–COX,* CAROL COUPLAND,* and RICHARD F. LOGAN‡

*Division of Primary Care, University of Nottingham, and ‡Division of Epidemiology and Public Health, University of Nottingham, Nottingham, United Kingdom

Background & Aims: **Several studies suggest that statins prevent some cancers, with one study finding a 47% reduction in colorectal cancer risk after ≥5 years of regular use.** ***Methods:*** **A nested case-control study was conducted within 454 general practices in the United Kingdom using the QRESEARCH database. Cases with colorectal cancer were diagnosed between 1995 and 2005. The effects of statins, nonsteroidal anti-inflammatory drugs, cyclooxygenase-2 inhibitors, and aspirin on colorectal cancer were estimated with conditional logistic regression adjusted for morbidity, smoking status, body mass index, and socioeconomic status.** ***Results:*** **We analyzed 5686 cases and 24,982 matched controls with ≥4 years of records. The adjusted odds ratio for colorectal cancer associated with any statin prescription was 0.93 (95% confidence interval: 0.83–1.04), with no trend in duration of use or number of prescriptions. For any nonsteroidal anti-inflammatory drug prescription the adjusted odds ratio was 0.94 (95% confidence interval: 0.88–1.00), with a significant decrease in risk with increasing number of prescriptions and an adjusted odds ratio of 0.76 (0.60–0.95) for ≥25 prescriptions. Prolonged use of cyclooxygenase-2 inhibitors was minimal, but for those receiving ≥25 prescriptions the adjusted odds ratio was 0.34 (0.14–0.85). Results were similar in the subset of participants with ≥8 years of records; the adjusted odds ratio for ≥61 months of statin prescriptions was 1.00 (0.67–1.48).** ***Conclusions:*** **In this large population-based case-control study prolonged use of nonsteroidal anti-inflammatory drug and cyclooxygenase-2 inhibitor was associated with a reduced colorectal cancer risk, but prolonged statin use was not.**

Colorectal cancer is the third most common cancer worldwide,[1] and effective chemoprevention agents would have important implications for public health. Laboratory data (mostly from studies in rodents) suggest that statins may be chemoprophylactic against various types of cancer, including colon[2] and breast cancers.[3,4] Statins appear to suppress the growth of cancer cells in vitro by causing the cells to pause in the G_1 phase of the mitotic cycle and by increasing cell death.[5] In contrast to the overwhelming evidence from randomized clinical trials for the beneficial effect of statins in vascular disease, their effects on the risk of cancer remains unclear. Greater clarity is obviously needed because statins are already being used for prolonged periods in large numbers of patients also at risk of colorectal cancer.[6]

Several clinical trials have reported on the risk of cancer in patients on statins, but generally the results were equivocal because of inadequate power. Three randomized trials involving statins reported no difference in the overall incidence of cancers,[7–9] whereas the PROspective Study of Pravastatin in the Elderly at Risk (PROSPER) trial, which included more elderly patients, reported a 46% increased risk of gastrointestinal cancer in the pravastatin arm.[10] A meta-analysis of the various cardiovascular trials performed to examine the impact of statins on cancer incidence was recently reported.[11] Of the 26 trials included, only 4 reported specific data on colorectal cancer incidence. Altogether there were 320 colorectal cancers reported, with no evidence of a reduced risk in the statin takers. However, only 2 of the 4 trials lasted >5 years, leaving open the possibility of some benefit from prolonged statin use.

Several observational cohort studies have also examined statin use and cancer risk but have been generally limited by small numbers of participants developing colorectal cancer and by short duration of statin exposure.[12–15] Nevertheless, a recent case-control study from Israel reported a 47% reduction in risk of colorectal cancer in patients reporting statin use of ≥5 years.[16]

We have undertaken a study to determine the risk of colorectal cancer in patients prescribed statins by using a

Abbreviations used in this paper: CI, confidence interval; COX-2, cyclooxygenase-2; GPRD, General Practice Research Database; NSAID, nonsteroidal anti-inflammatory drug; OR, odds ratio.

0016-5085/07/$32.00
doi:10.1053/j.gastro.2007.05.023

large population-based general practice database. In addition, we included in the protocol an analysis to determine the risk of colorectal cancer in patients prescribed nonsteroidal anti-inflammatory drugs (both traditional [NSAIDs] and cyclooxygenase-2 [COX-2] inhibitors). This inclusion was to substantiate previous findings from British primary care for traditional NSAIDs[17,18] and to offer new data on COX-2 inhibitors in the light of recent colorectal adenoma prevention trials that were terminated because of safety concerns.[19,20]

Materials and Methods

Study Population and Data Source

We conducted the study using general practices in the United Kingdom contributing to the QRESEARCH database (http://www.qresearch.org). This is a new clinical database containing the records of almost 8 million patients ever registered with 454 practices during the past 16 years. The information recorded in the database includes patient demographics (year of birth, sex, and socioeconomic data associated with postcode area), characteristics (height, weight, smoking status), symptoms, clinical diagnoses (Read codes), consultations, referrals, prescribed medication, and results of investigations. Version 8 of the QRESEARCH database was used for this analysis.

The QRESEARCH database has been validated by comparing birth rates, death rates, consultation rates, and prevalence and mortality rates with other data sources, including the General Household Survey and the General Practice Research Database (GPRD).[21] Correspondence is good for all of these measures (results available on request), although in some instances QRESEARCH prevalence figures of chronic diseases such as diabetes, hypertension, and stroke are marginally higher than less recent data.[22] The age-sex structure of the QRESEARCH population is similar to that reported in the United Kingdom 2001 census. We have also compared practices taking part in regional research networks on these and other measures and found a good correspondence.[23] Detailed analyses have shown good levels of completeness and consistency.[24] The database has been used for studies that investigate effects of NSAIDs[25,26] and statin.[27] The diagnosis of cancer in primary care databases was found to be sufficiently reliable to allow analysis of cancer risk in relation to the prescribing of calcium channel blockers.[28]

Cohort Definition

Our study period for this analysis was the 10 years between January 1, 1995 and July 31, 2005 (the date of the most recent download available at the time of the study). We identified an open cohort of patients registered on or after January 1, 1995. Our left censor date was the latest of the patients' registration date or January 1, 1995. Our right censor date was the earliest of the dates on which they developed colorectal cancer, died, left the practice, or the study period ended.

Cases of colorectal cancer were identified on the basis of a first-time computer-recorded diagnosis of colorectal cancer during the 10-year study period. Patients with a recorded malignancy before the study period were excluded.

We determined the crude incidence rate of colorectal cancer for men and women and compared this to national incidence data as part of our validation processes.

Cancer Cases and Controls

We assembled matched case-control sets in which cases were all patients with an incident colorectal cancer during the 10-year study period. With the use of incidence density sampling, we matched up to 5 controls to each case by age (within a year), calendar time, sex, and practice. All controls were alive and registered with the practice and free of colorectal cancer at the time their matched case was diagnosed. We derived an index date for each control which corresponded to the first recorded date of diagnosis of colorectal cancer in the matched case.

Assessment of Exposure

We restricted the main statistical analyses to subjects with $\geq$4 years of records available before their index date. We reviewed the medical history and extracted data on prescribed medications before the index date for each set of cases and controls.

For the analyses, a patient was assumed to be exposed to a drug if the patient had received $\geq$1 prescription for that drug in the 13 to 48 months before his or her index date. We ignored prescriptions issued in the 12 months immediately preceding the date of diagnosis of colorectal cancer or the equivalent date in controls. This was done to minimize issues of reverse causality; for example, patients with prodromal symptoms could consult in the year before diagnosis and have a serum cholesterol measurement as part of a general screening procedure and hence be prescribed statins as a result.

We grouped the drugs as follows: statins (atorvastatin, cerivastatin, fluvastatin, pravastatin, and simvastatin), NSAIDs (ibuprofen, diclofenac, naproxen, and other nonselective NSAIDs), COX-2 inhibitors (meloxicam, celecoxib, rofecoxib, etoricoxib, etodolac, valdecoxib), and aspirin. Apart from ibuprofen and aspirin, none of these drugs was available without prescription during the study period.

For each statin, we identified each prescription issued during the 13–48 months before the index date, then extracted dose and duration in days for each statin prescription. We estimated the cumulative duration in days for all statin prescriptions during the 13- to 48-month interval and converted the duration to months, assuming

that 12 months were equivalent to 365 days. Because prescribing of the other drugs under analysis was less continuous and recommended dosage was more variable, we restricted our calculations of duration of exposure to number of prescriptions for these drugs. General practitioners in the United Kingdom issue patients with sufficient drugs to last ≥1 calendar month, so one prescription is approximately equivalent to 1 month of treatment. We grouped the number of prescriptions in the last 13–48 months as only 1 prescription, 2–12, 13–24, and ≥25 prescriptions.

For our primary exposure of interest (statins), we also conducted analyses for each individual type of statin medication. For the analyses of interactions between drugs, we considered statins to have been prescribed at the same time as NSAIDs or COX-2 inhibitors if the drugs were prescribed within 90 days of each other.

Confounding Variables

We considered smoking and obesity to be possible confounding factors for colorectal cancer.[1] We also took account of the following morbidities if they were diagnosed ≥13 months before the index date: ulcerative colitis, diabetes, ischemic heart disease with and without a history of myocardial infarction, hypertension, stroke, rheumatoid arthritis, and osteoarthritis. We adjusted for socioeconomic status with the Townsend deprivation score based on 2001 postcode-related census data. This is an area-level composite score based on unemployment, overcrowding, lack of home ownership, and lack of car ownership, and it is strongly related to morbidity.[29] Higher scores indicate greater levels of material deprivation.

Statistical Analysis

We estimated the odds ratio of colorectal cancer for each drug group using conditional logistic regression analysis for individually matched case-control studies. The odds ratios (ORs) and 95% confidence intervals (CIs) were adjusted for possible confounding effects of morbidity (as listed previously), smoking status (smoker, not smoker, not recorded), body mass index (calculated as weight in kilograms divided by the square of height in meters [kg/m^2]; <25, 25 to 29.9, ≥30, not recorded), socioeconomic status (in fifths), and use of the other drug groups (statins, any traditional NSAID, any COX-2 inhibitor, and aspirin).

We undertook tests for trend across the number of prescriptions and the duration of statin use, using ordinal variables and examining the significance of the coefficients with adjusted Wald's tests. We tested for interactions between statins and NSAIDs and statins and COX-2 inhibitors. We preselected a P value of .01 as indicating statistical significance, to take account of the size of the dataset and the potential for multiple comparisons. All P values are two-sided.

Additional Analyses

We repeated the analysis, restricting it to patients with ≥8 years of complete prescribing data. In this analysis we grouped the number of prescriptions in the past 13–96 months as only 1 prescription; 2–12; 13–24; 25–36; 37–48; ≥49. The duration of statin use was grouped as none, ≤12 months, 13–24 months, 25–36 months, 37–48 months, 49–60 months, and ≥61. We also conducted analyses restricted to patients with complete data for smoking status, body mass index, and deprivation.

The study was approved by the Trent Multicentre Ethics committee and the QRESEARCH Scientific Advisory Board. The study had no external funding and was conducted independently of the pharmaceutical industry.

Results

The total number of patients included in the cohort was 1,896,944 patients registered within a total of 454 practices. We identified 9694 incident cases of colorectal cancer between January 1995 and July 2005 arising from 8,823,664 person-years of observation. The crude incidence rate of colorectal cancer was 49.8 per 100,000 person years (56.1 in men and 43.6 in women). In comparison colorectal cancer in 2003 in the United Kingdom has been reported as 62.3 per 100,000 in men and 49.5 per 100,000 in women.[30] Of the 9694 cases of incident colorectal cancer, 5686 cases, matched to 24,982 controls, had a minimum of 4 years of registration with that general practice.

Baseline Characteristics

Table 1 shows the baseline characteristics of cases with colorectal cancer and their matched controls: 3181 of the colorectal cancer patients were men (55.9%); and their median age at diagnosis was 72 years (interquartile range: 64–79). Of the cases, 3460 (60.9%) had colon cancer and 2226 (39.1%) had rectal cancer.

An average of 4.4 controls was identified for each case. The median number of months of prior data for both case and control groups was 88 months (interquartile range: 66–117). Cases and controls had similar patterns of comorbidity except for a higher prevalence of diabetes (8.7% cases vs 6.8% controls) and colitis (1% cases vs 0.6% controls) in cases and a lower prevalence of rheumatoid arthritis (0.9% cases vs 1.4% controls). The baseline characteristics for the subset of 2425 cases and 9706 matched controls with ≥8 years of medical records were similar to the sample with ≥4 years of records (data available from the authors).

Use of Statins

Table 2 shows the frequencies and ORs for use of statins in cases and controls, by duration of prescriptions in months, and the number of prescriptions in the previous 13–48 months. Ninety-five percent of cases and of

Table 1. Characteristics of Cases and Matched Controls with ≥4 Years of Records and Odds Ratios for the Variables

Characteristics	Cases n = 5686	Controls n = 24,982	Odds ratios (95% confidence interval)[a]
Men, n (%)	3181 (55.9)	14,014 (56.1)	
Females, n (%)	2505 (44.1)	10,968 (43.9)	
Age group			
<55 y, n (%)	522 (9.2)	2128 (8.5)	
55–64 y, n (%)	1007 (17.7)	4412 (17.7)	
65–74 y, n (%)	1818 (32.0)	8103 (32.4)	
75–84 y, n (%)	1867 (32.8)	8340 (33.4)	
≥85 y, n (%)	472 (8.3)	1999 (8.0)	
Deprivation			
Townsend score, median (interquartile range)	−1.26 (−3.05, 1.57)	−1.43 (−3.16, 1.46)	
Townsend quintile 1 most affluent, n (%)	1302 (22.9)	5945 (23.8)	1.00
Townsend quintile 2, n (%)	1208 (21.3)	5495 (22.0)	1.01 (0.92–1.10)
Townsend quintile 3, n (%)	1163 (20.5)	4788 (19.2)	1.11 (1.02–1.22)
Townsend quintile 4, n (%)	946 (16.6)	4070 (16.3)	1.07 (0.97–1.18)
Townsend quintile 5 most deprived, n (%)	899 (15.8)	3753 (15.0)	1.10 (0.98–1.23)
Townsend quintile missing, n (%)	168 (3.0)	931 (3.7)	
Body mass index			
<25 kg/m^2, n (%)	1686 (29.7)	6928 (27.7)	1.00
25–29.9 kg/m^2, n (%)	1835 (32.3)	7687 (30.8)	0.99 (0.92–1.07)
≥30 kg/m^2, n (%)	839 (14.8)	3581 (14.3)	0.95 (0.87–1.05)
Not recorded, n (%)	1326 (23.3)	6786 (27.2)	
Smoking status			
Nonsmoker, n (%)	3845 (67.6)	16,212 (64.9)	1.00
Smoker, n (%)	985 (17.3)	4060 (16.3)	1.02 (0.95–1.11)
Smoking status not recorded, n (%)	856 (15.1)	4710 (18.9)	
Morbidity			
Ischemic heart disease			
No myocardial infarction, n (%)	488 (8.6)	2067 (8.3)	1.01 (0.90–1.14)
Myocardial infarction, n (%)	271 (4.8)	1313 (5.3)	0.90 (0.78–1.05)
Diabetes, n (%)	493 (8.7)	1697 (6.8)	1.26 (1.13–1.41)
Hypertension, n (%)	1716 (30.2)	7312 (29.3)	1.01 (0.94–1.08)
Osteoarthritis, n (%)	684 (12.0)	3033 (12.1)	0.99 (0.90–1.09)
Colitis, n (%)	57 (1.0)	145 (0.6)	1.70 (1.25–2.32)
Rheumatoid arthritis, n (%)	52 (0.9)	353 (1.4)	0.65 (0.48–0.87)
Stroke, n (%)	306 (5.4)	1292 (5.2)	1.04 (0.91–1.19)

[a]Odds ratios for the model which includes smoking, obesity, deprivation, morbidity (diabetes, ischemic heart disease, hypertension, stroke, rheumatoid arthritis, and osteoarthritis) and use of any statin, any cyclooxygenase-2 inhibitor, any nonsteroidal anti-inflammatory drug, aspirin.

controls who were prescribed statins for >24 months in this period continued to use them in the 12 months before the index date. The majority of statin use was continuous: 90% of cases and controls who had >1 statin prescription in the previous 13–48 months had no break in prescribing of >3 months and 96% of cases and controls had no break of >6 months. No statistically significant trends were observed in the adjusted ORs associated with either the duration or the number of statin prescriptions (Figure 1). Although the upper 95% CI for a single statin prescription is less than unity, the *P* value of .04 is not considered statistically significant.

Table 2 also shows the frequencies and ORs for use of statins in the subgroup with ≥8 years of records available. The adjusted ORs for any use of statins was 0.94 (95% CI: 0.79–1.11), and no significant associations were observed with the duration of use or the number of prescriptions.

The statins most frequently prescribed were atorvastatin (4.4% of cases and 3.8% of controls) and simvastatin (5.0% of cases and 5.7% of controls) with <1% of cases and controls having prescriptions for other statins (cerivastatin, fluvastatin, and pravastatin) (Table 3).

We found some variation in the ORs for colorectal cancer associated with individual statins. In the unadjusted analysis, any use of atorvastatin or cerivastatin was associated with increased ORs for colorectal cancer, although these did not reach the 0.01 significance level before or after adjustment. Although any use of simvastatin, after adjustment for confounders, including use of other statins, was associated with a 17% decrease in cancer risk (adjusted OR: 0.83; 95% CI: 0.72–0.96; $P = .013$), no significant trend was observed with the number of simvastatin prescriptions.

Use of NSAIDs, COX-2 Inhibitors, and Aspirin

Table 4 show the frequencies and the ORs for COX-2 inhibitors, traditional NSAIDs, and aspirin use by the number of prescriptions for these medications in the

Table 2. Use of Statins Before Index Date in Cases and Controls

	Cases n (%)	Controls n (%)	Unadjusted odds ratio (95% CI)	Adjusted odds ratio[a] (95% CI)	P[a]
Use in 13–48 mo[b]					
Any statin	538 (9.5)	2424 (9.7)	0.99 (0.89–1.09)	0.93 (0.83–1.04)	.22
Duration in 13–48 mo[b]					
None	5148 (90.5)	22,558 (90.3)	1.00	1.00	.69[c]
1–12 mo	183 (3.2)	911 (3.6)	0.90 (0.76–1.06)	0.84 (0.71–1.00)	
13–24 mo	122 (2.1)	526 (2.1)	1.04 (0.85–1.27)	0.99 (0.80–1.22)	
25+ mo	233 (4.1)	987 (4.0)	1.04 (0.90–1.21)	0.99 (0.84–1.16)	
No. prescriptions in 13–48 mo[b]					
None	5148 (90.5)	22,558 (90.3)	1.00	1.00	.99[c]
1	32 (0.6)	189 (0.8)	0.74 (0.51–1.08)	0.67 (0.46–0.98)	
2–12	206 (3.6)	1007 (4.0)	0.91 (0.78–1.07)	0.85 (0.72–1.01)	
13–24	170 (3.0)	726 (2.9)	1.05 (0.88–1.24)	1.01 (0.84–1.21)	
25+	130 (2.3)	502 (2.0)	1.15 (0.94–1.40)	1.13 (0.91–1.41)	
Use in 13–96 mo[d]					
Any statin	302 (12.5)	1220 (12.6)	1.01 (0.87–1.16)	0.94 (0.79–1.11)	.44
Duration in 13–96 mo[d]					
None	2123 (87.5)	8486 (87.4)	1.00	1.00	.44[c]
1–12 mo	115 (4.7)	440 (4.5)	1.06 (0.85–1.31)	0.98 (0.78–1.23)	
13–24 mo	60 (2.5)	259 (2.7)	0.96 (0.72–1.28)	0.90 (0.67–1.22)	
25–36 mo	43 (1.8)	167 (1.7)	1.05 (0.75–1.48)	0.97 (0.68–1.39)	
37–48 mo	27 (1.1)	124 (1.3)	0.87 (0.57–1.34)	0.82 (0.53–1.27)	
49–60 mo	22 (0.9)	98 (1.0)	0.90 (0.56–1.44)	0.83 (0.51–1.35)	
61+ mo	35 (1.4)	132 (1.4)	1.07 (0.73–1.57)	1.00 (0.67–1.48)	
No. of prescriptions in 13–96 mo[d]					
None	2123 (87.5)	8486 (87.4)	1.00	1.00	.63[c]
1	23 (0.9)	97 (1.0)	0.99 (0.63–1.57)	0.87 (0.55–1.40)	
2–12	122 (5.0)	469 (4.8)	1.06 (0.86–1.31)	1.00 (0.80–1.25)	
13–24	58 (2.4)	271 (2.8)	0.87 (0.65–1.16)	0.81 (0.59–1.10)	
25–36	46 (1.9)	176 (1.8)	1.05 (0.75–1.48)	0.97 (0.68–1.38)	
37–48	22 (0.9)	96 (1.0)	0.95 (0.60–1.52)	0.88 (0.55–1.43)	
49+	31 (1.3)	111 (1.1)	1.11 (0.73–1.67)	1.11 (0.72–1.72)	

[a]Adjusted for smoking, obesity, deprivation, morbidity (diabetes, ischemic heart disease, hypertension, stroke, colitis, rheumatoid arthritis, and osteoarthritis), use of the other medications (number of prescriptions).
[b]Cases (n = 5686) and controls (n = 24,982).
[c]Trend test.
[d]Cases (n = 2425) and controls (n = 9706).

previous 13–48 and 13–96 months. Patients receiving ≥25 prescriptions for traditional NSAIDs in the past 13–48 months had a lower risk of colorectal cancer than patients not prescribed NSAIDs (adjusted OR: 0.76; 95% CI: 0.60–0.95), and the test for trend was highly significant (Table 4; Figure 1). In this group receiving ≥25 NSAID prescriptions in the past 13–48 months 77% of cases and 80% of controls continued to use these drugs in the 12 months before the index date. A significant decrease in risk was observed in patients who had ≥25 prescriptions of COX-2 inhibitors in the past 13–48 months (adjusted OR: 0.34; 95% CI: 0.14–0.85) compared with patients not prescribed COX-2 inhibitors, although the test for trend was not statistically significant. In the group receiving ≥25 COX-2 inhibitor prescriptions in the 13–49 months before the index date 100% of cases and 91% of controls also received COX-2 inhibitor prescriptions in the 12 months before the index date. For aspirin use, for which 70% of prescriptions were for ≤75 mg/day, about a 10% reduction was observed in risk associated with ≥13 prescriptions in the past 13–48 months, but the trend was not significant. In the group with ≥25 aspirin prescriptions, most continued to take aspirin in the year before the index date (95% of cases and 90% of controls).

The adjusted OR for use of any traditional NSAID in the past 13–96 months was 0.91 (95% CI: 0.83–1.00), with longer use showing a more protective effect (*P* for trend = 0.001) (Figure 2). For aspirin use about a 10% to 15% reduction in risk associated with ≥37 prescriptions was observed in the past 13–96 months, but the trend was not significant. For this analysis it is important to note that COX-2 inhibitors were not in use for the first 5 years of the study period. No significant interactions were observed between any of the drug groups (statins, NSAIDs, and COX-2 inhibitors).

We repeated all of the above analyses separately for colon and rectal cancers and found similar results for the separate diagnoses. We also repeated the analyses, restricting them to cases and controls with complete data on

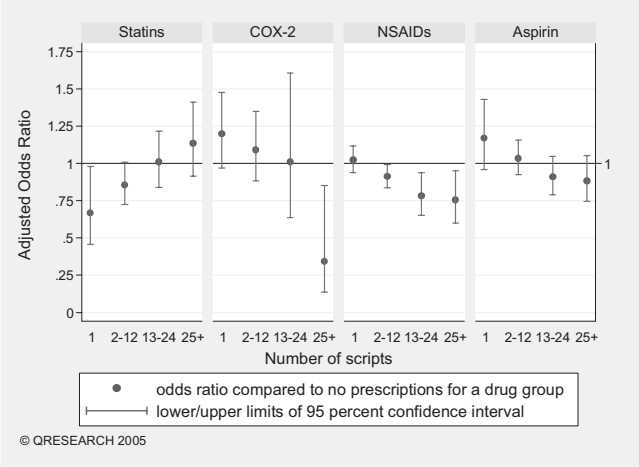

Figure 1. Number of prescriptions for the different drugs in 13–48 months before the index date and adjusted odds ratios for colorectal cancer.

smoking, body mass index, and deprivation (71% of cases and 54% of controls) and restricting them to patients aged ≥65 years, and obtained similar results for all groups of drugs and individual types of statins. Finally, to check on the possibility of confounding by hyperlipidemia as the indication for statin use, we included a variable for this diagnosis (recorded in 4.9% of cases and 4.8% of controls) and found our results were essentially unchanged.

Discussion

This is a large population-based study designed to determine the association between the use of statins and development of colorectal cancer. Although we were able to confirm previous protective associations between colorectal cancer and traditional NSAIDs, we were unable to confirm the large reduction in colorectal cancer risk with prolonged statin use reported in the recent case-control study from Israel.[16] However, equally, it also provides reassurance that statins as a class do not increase the risk of colon cancer, a concern raised within the PROSPER pravastatin trial.[10]

In contrast to our findings on statin use, prolonged use of NSAIDs was associated with a ≥25% reduction in colorectal cancer risk, similar to that found in a previous case-control study of colorectal cancer using another British primary care database, GPRD.[18] In that study the adjusted OR for colorectal cancer among patients using traditional NSAIDs for >2 years was 0.66 (95% CI: 0.40–0.80), which is similar to our value of 0.76 (95% CI: 0.60–0.95) for ≥25 prescriptions of NSAIDs in the past 13–48 months. Other established risk factors (such as diabetes and ulcerative colitis) also showed a positive

Table 3. Use of Individual Statins in 13–48 Months Before the Index Date in 5686 Cases and 24,982 Controls

	Cases n (%)	Controls n (%)	Unadjusted odds ratio (95% CI)	Adjusted odds ratio[a] (95% CI)	*P*[a]
Atorvastatin	250 (4.4)	954 (3.8)	1.19 (1.02–1.38)	1.11 (0.95–1.30)	.17
Simvastatin	287 (5.0)	1420 (5.7)	0.88 (0.77–1.01)	0.83 (0.72–0.96)	.013
Pravastatin	27 (0.5)	136 (0.5)	0.90 (0.59–1.37)	0.84 (0.55–1.28)	.41
Fluvastatin	44 (0.8)	150 (0.6)	1.35 (0.95–1.91)	1.21 (0.85–1.74)	.29
Cerivastatin	56 (1.0)	179 (0.7)	1.40 (1.03–1.91)	1.34 (0.97–1.86)	.07
Duration of atorvastatin use					
None	5436 (95.6)	24,028 (96.2)	1.00	1.00	.14[b]
1–12 mo	111 (1.6)	452 (1.8)	1.12 (0.90–1.38)	1.07 (0.85–1.33)	
13–24 mo	60 (1.1)	223 (0.9)	1.20 (0.90–1.61)	1.13 (0.84–1.53)	
25+ mo	79 (1.4)	279 (1.1)	1.29 (1.00–1.67)	1.20 (0.92–1.56)	
Duration of simvastatin use					
None	5399 (95.0)	23,562 (94.3)	1.00	1.00	.07[b]
1–12 mo	119 (2.1)	586 (2.4)	0.89 (0.73–1.09)	0.82 (0.66–1.01)	
13–24 mo	56 (1.0)	317 (1.3)	0.78 (0.58–1.04)	0.73 (0.54–0.98)	
25+ mo	112 (2.0)	517 (2.1)	0.94 (0.76–1.16)	0.92 (0.74–1.15)	
No. of prescriptions for atorvastatin					
None	5436 (95.6)	24,028 (96.2)	1.00	1.00	.10[b]
1	16 (0.3)	99 (0.4)	0.72 (0.42–1.22)	0.66 (0.39–1.14)	
2–12	122 (2.1)	457 (1.8)	1.21 (0.98–1.49)	1.14 (0.92–1.42)	
13–24	74 (1.3)	248 (1.0)	1.36 (1.04–1.77)	1.28 (0.97–1.69)	
25+	38 (0.7)	150 (0.6)	1.16 (0.80–1.67)	1.13 (0.78–1.65)	
No. of prescriptions for simvastatin					
None	5399 (95.0)	23,562 (94.3)	1.00	1.00	.11[b]
1	24 (0.4)	138 (0.6)	0.75 (0.48–1.15)	0.67 (0.43–1.04)	
2–12	127 (2.2)	631 (2.5)	0.89 (0.73–1.08)	0.82 (0.67–1.01)	
13–24	70 (1.2)	415 (1.7)	0.74 (0.57–0.96)	0.73 (0.56–0.95)	
25+	66 (1.2)	236 (0.9)	1.20 (0.91–1.59)	1.22 (0.91–1.64)	

[a]Adjusted for smoking, obesity, deprivation, morbidity (diabetes, ischemic heart disease, hypertension, stroke, colitis, rheumatoid arthritis, and osteoarthritis), use of nonsteroidal anti-inflammatory drug, cyclooxygenase-2, aspirin, and the other statins.
[b]Trend test.

Table 4. Use of Anti-Inflammatory Medication Before the Index Date

	Cases (%)	Controls (%)	Unadjusted odds ratio (95% CI)	Adjusted odds ratio[a] (95% CI)	P[a]
Use in 13–48 mo[b]					
Any COX-2	263 (4.6)	1101 (4.4)	1.05 (0.91–1.21)	1.07 (0.92–1.24)	.37
Any NSAID	1871 (32.9)	8460 (33.9)	0.96 (0.90–1.02)	0.94 (0.88–1.00)	.048
Aspirin	1226 (21.6)	5369 (21.5)	1.01 (0.94–1.09)	0.99 (0.90–1.08)	.79
No. of COX-2 prescriptions in 13–48 mo[b]					
None	5423 (95.4)	23,881 (95.6)	1.00	1.00	.88[c]
1	118 (2.1)	444 (1.8)	1.18 (0.96–1.45)	1.20 (0.97–1.48)	
2–12	117 (2.1)	487 (1.9)	1.06 (0.86–1.31)	1.09 (0.88–1.35)	
13–24	23 (0.4)	103 (0.4)	0.93 (0.59–1.48)	1.01 (0.64–1.61)	
25+	5 (0.1)	67 (0.3)	0.33 (0.13–0.83)	0.34 (0.14–0.85)	
No. of NSAID prescriptions in 13–48 mo[b]					
None	3815 (67.1)	16,522 (66.1)	1.00	1.00	.001[c]
1	765 (13.5)	3130 (12.5)	1.06 (0.97–1.15)	1.02 (0.94–1.12)	
2–12	864 (15.2)	3992 (16.0)	0.94 (0.87–1.02)	0.91 (0.84–0.99)	
13–24	151 (2.7)	822 (3.3)	0.80 (0.67–0.96)	0.78 (0.65–0.94)	
25+	91 (1.6)	516 (2.1)	0.76 (0.61–0.96)	0.76 (0.60–0.95)	
No. of aspirin prescriptions in 13–48 mo[b]					
None	4460 (78.4)	19,613 (78.5)	1.00	1.00	.19[c]
1	132 (2.3)	494 (2.0)	1.17 (0.96–1.42)	1.17 (0.96–1.43)	
2–12	541 (9.5)	2281 (9.1)	1.06 (0.96–1.17)	1.03 (0.92–1.16)	
13–24	334 (5.9)	1563 (6.3)	0.95 (0.83–1.07)	0.91 (0.79–1.05)	
25+	219 (3.9)	1031 (4.1)	0.94 (0.80–1.10)	0.88 (0.74–1.05)	
Use of medications in 13–96 mo[d]					
Any NSAID	1211 (49.9)	4989 (51.4)	0.94 (0.86–1.03)	0.91 (0.83–1.00)	.06
Aspirin	636 (26.2)	2529 (26.1)	1.03 (0.92–1.14)	0.98 (0.86–1.12)	.77
No. of NSAID prescriptions in 13–96 mo[d]					
None	1214 (50.1)	4717 (48.6)	1.00	1.00	.001[c]
1	377 (15.5)	1456 (15.0)	1.00 (0.88–1.14)	0.97 (0.85–1.10)	
2–12	639 (26.4)	2513 (25.9)	0.98 (0.88–1.09)	0.94 (0.84–1.05)	
13–24	93 (3.8)	422 (4.3)	0.86 (0.68–1.09)	0.83 (0.65–1.06)	
25–36	46 (1.9)	243 (2.5)	0.74 (0.54–1.02)	0.73 (0.52–1.01)	
37–48	21 (0.9)	159 (1.6)	0.50 (0.31–0.79)	0.48 (0.30–0.77)	
49+	35 (1.4)	196 (2.0)	0.69 (0.48–1.00)	0.69 (0.48–1.00)	
No. of aspirin prescriptions in 13–96 mo[d]					
None	1789 (73.8)	7177 (73.9)	1.00	1.00	.47[c]
1	72 (3.0)	273 (2.8)	1.07 (0.82–1.40)	1.03 (0.79–1.36)	
2–12	190 (7.8)	782 (8.1)	1.01 (0.85–1.20)	0.95 (0.79–1.14)	
13–24	149 (6.1)	539 (5.6)	1.11 (0.91–1.34)	1.06 (0.86–1.31)	
25–36	94 (3.9)	361 (3.7)	1.07 (0.84–1.36)	1.01 (0.78–1.32)	
37–48	54 (2.2)	237 (2.4)	0.95 (0.70–1.29)	0.91 (0.66–1.26)	
49+	77 (3.2)	337 (3.5)	0.90 (0.69–1.17)	0.85 (0.63–1.14)	

[a]Adjusted for smoking, obesity, deprivation, morbidity (diabetes, ischemic heart disease, hypertension, stroke, colitis, rheumatoid arthritis, and osteoarthritis), use of the other medications (number of prescriptions).
[b]Cases (n = 5686) and controls (n = 24,982).
[c]Trend test.
[d]Cases (n = 2425) and controls (n = 9706).

association with colorectal cancer in our study as reported elsewhere.[1] Like the GPRD study, we found around a 10% reduction in risk associated with prolonged aspirin use (≥13 prescriptions in the past 13–48 months), which was not statistically significant. Other studies have suggested that the benefit as a result of aspirin may take more than a decade to accrue[31] and requires a dose >75 mg daily, which is likely to explain our findings in this context.

We also found some evidence that prolonged use of COX-2 inhibitors was associated with a significantly reduced risk of colorectal cancer; we found a 66% reduction in risk in patients who had ≥25 prescriptions than patients who had not been prescribed COX-2 inhibitors. This is of interest in view of recent trials with celecoxib which reported a ≥50% reduction in the occurrence of advanced adenoma and the overexpression of COX-2 that was shown in colorectal cancers.[20,32] However, examining the risk of colorectal cancer in patients taking COX-2 inhibitors was a secondary aim of our study, overall usage was low, and the test for trend was not significant, so this finding needs to be interpreted with caution.

Our study has several strengths. It is substantially larger and has greater statistical power than previous

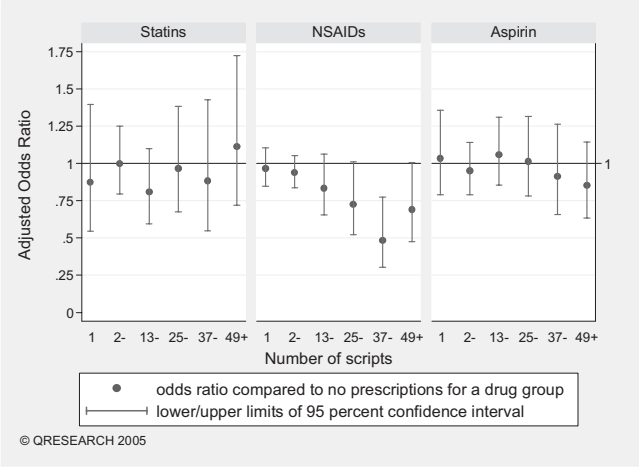

Figure 2. Number of prescriptions for the different drugs in 13–96 months before the index date and adjusted odds ratios for colorectal cancer.

studies.[12–16,33] Because it is based on computer-recorded prescribing and morbidity data collected prospectively, we were able to include all patients (including those who had died) rather than being restricted to a survivor volunteer population as in the study of Poynter et al[16] in which only two thirds of eligible cases and half of the eligible controls were included.

Matching of controls to cases on age, sex, calendar time, and practice removed confounding by these factors. Unlike the Israeli study[16] and a recent Massachusetts study,[34] recall bias for the type and duration of statin and other drug use is not an issue because information about the patient and drugs prescribed was recorded on computer before the diagnosis of cancer was made, and so the information was unaffected by the cancer diagnosis itself. Any bias from misclassification is likely to be minimal because recording of clinical diagnoses and prescribed medication in general practice was shown to have high levels of accuracy and completeness.[35] In addition, statins were available only on prescription throughout the study period. The similar results for ibuprofen and aspirin use in patients aged ≥65 years, who are entitled to free prescribed medications and so unlikely to buy them over the counter, suggest that misclassification of use of these medications because of over-the-counter purchase is not an explanation for our findings.

Our study also had some limitations. Information on certain risk factors for colorectal cancer, such as sedentary lifestyle, family history, and diet,[1] are not recorded on the database and could not be included in the analysis. No information was available on cancer stage, and information on how the cancer was treated was incomplete. Other factors such as body weight, alcohol intake, and smoking status are less consistently recorded, because the general practitioner either does not ask or does not record the relevant information; hence, there may be some misclassification for these factors. Some confounding may remain if these factors are also associated with statin use. Nevertheless, in the study by Poynter et al,[16] which was able to adjust for sports participation, a family history of colorectal cancer, and level of vegetable consumption, the effect of adjustment was small.

Our incidence rates were slightly lower than national figures, suggesting possible under-ascertainment of cases. The under-ascertainment is likely to be due to some colorectal cancers only being registered at the time of death which may go unrecorded in the general practitioner records.[36] However, making the assumption that the underrecording rate is ≤10%, <16 of the 24,982 sampled controls are likely to be unrecorded cases, a level of under-ascertainment unlikely to have an influence on our findings. It is also possible that statin users might be more likely to have colorectal cancer detected as an indirect consequence of more frequent practice attendance. Although ignoring statin prescribing in the 12 months before the diagnosis date will have reduced this bias, it will not entirely eliminate the possibility of detection bias.

Although our data contain detailed information on drug prescriptions, this may not reflect actual use. However, there is no reason to think that any nonadherence would systematically differ between cases and controls. Even though this is the largest study of its kind, there were only a relatively small number of participants (1.4% of cases and controls) with ≥8 years of records and prolonged exposure to statins. Thus, the 95% CI for the most prolonged statin use (61+months) is consistent with both a 33% reduction in colorectal cancer risk as well as a 48% increase. Nevertheless, there was no hint of any dose-response relationship with statin use in this subset or in the full dataset. We also cannot exclude the possibility that protection from colorectal cancer is confined to a particular statin. In this regard the data for simvastatin could be interpreted as hinting at some reduction in cancer risk. However, we had no prior hypothesis here, and it is notable that there was no indication of a protective effect specific to simvastatin in previous case-control studies.[16,34]

In summary, we have conducted a large population-based case-control study that examined the effect of statins on the risk of colorectal cancer and found that, although prolonged NSAID and COX-2 inhibitor use are associated with reduced colorectal cancer risk, prolonged statin use is not.

References

1. Weitz J, Koch M, Debus J, Hohler T, Galle PR, Buchler MW. Colorectal cancer. Lancet 2005;365:153–165.
2. Carlberg M, Dricu A, Blegen H, Wang M, Hjertman M, Zickert P, Hoog A, Larsson O. Mevalonic acid is limiting for N-linked glycosylation and translocation of the insulin-like growth factor-1 receptor to the cell surface. Evidence for a new link between 3-hydroxy-3-methylglutaryl-coenzyme a reductase and cell growth. J Biol Chem 1996;271:17453–17462.

3. Addeo R, Altucci L, Battista T, Bonapace IM, Cancemi M, Cicatiello L, Germano D, Pacilio C, Salzano S, Bresciani F, Weisz A. Stimulation of human breast cancer MCF-7 cells with estrogen prevents cell cycle arrest by HMG-CoA reductase inhibitors. Biochem Biophys Res Commun 1996;220:864–870.
4. Inano H, Suzuki K, Onoda M, Wakabayashi K. Anti-carcinogenic activity of simvastatin during the promotion phase of radiation-induced mammary tumorigenesis of rats. Carcinogenesis 1997; 18:1723–1727.
5. Keyomarsi K, Sandoval L, Band V, Pardee AB. Synchronization of tumor and normal cells from G1 to multiple cell cycles by lovastatin. Cancer Res 1991;51:3602–3609.
6. Agarwal B, Bhendwal S, Halmos B, Moss SF, Ramey WG, Holt PR. Lovastatin augments apoptosis induced by chemotherapeutic agents in colon cancer cells. ClinCancer Res 1999;5:2223–2229.
7. Downs JR, Clearfield M, Weis S, Whitney E, Shipiro DR, Beere PA, Langendorfer A, Stein EA, Kruyer W, Gotto AM Jr. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels. JAMA 1998;279:1349–1357.
8. The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in moderately hypercholesterolemic, hypertensive patients randomized to pravastatin vs usual care: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT-LLT). JAMA 2002;288:2998–3007.
9. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high risk individuals: a randomised placebo-controlled trial. Lancet 2002;360:7–22.
10. Shepherd J, Blauw G, Murphy M, Bollen E, Buckley B, Cobbe S, Ford I, Gaw A, Hyland M, Jukema JW, Kamper AM, Macfarlane PW, Meinders AE, Norrie J, Packard CJ, Perry IJ, Stott DJ, Sweeney BJ, Twomey C, Westendorp RG; PROSPER study group. PROspective Study of Pravastatin in the Elderly at Risk. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. Lancet 2002;360:1623–1630.
11. Dale K, Coleman C, Hentyan N, Kluger J, White C. Statins and cancer risk–a meta-analysis. JAMA 2006;295:74–80.
12. Blais L, Desgagne A, LeLorier J. 3-Hydroxy-3-methylglutaryl coenzyme A reductase inhibitors and the risk of cancer. Arch Intern Med 2000;160:2363–2368.
13. Fris S, Poulsen AH, Johnson SP, McLaughlin JK, Fryzek JP, Dalton SO, Sorensen HT, Olsen JH. Cancer risk among statin users: A population-based cohort study. Int J Cancer 2005;114: 643–647.
14. Setoguchi S, Glynn RJ, Avorn J, Mogun H, Schneeweiss S. Statins and the risk of lung, breast, and colorectal cancer in the elderly. Circulation 2007;115:27–33.
15. Jacobs EJ, Rodriguez C, Brady KA, Connell CJ, Thun MJ, Calle EE. Cholesterol-lowering drugs and colorectal cancer incidence in a large United States cohort. J Natl Cancer Inst 2006;98:69–72.
16. Poynter JN, Gruber SB, Higgins PDR, Almog R, Bonner JD, Rennert HS, Low M, Greenson JK, Rennert G. Statins and risk of colorectal cancer. N Engl J Med 2005;352:2184–2192.
17. Thun M, Henley S, Patrono C. Nonsteroidal anti-inflammatory drugs as anticancer agents: mechanistic, pharmacological and clinical issues. J Natl Cancer Inst 2002;94:252–266.
18. Garcia Rodriguez LA, Huerto-Alvarez C. Reduced risk of colorectal cancer among long-term users of aspirin and nonaspirin nonsteroidal anti-inflammatory drugs. Epidemiology 2001;12:88–93.
19. Bresalier RS, Sandler RS, Quan H, Bolognese JA, Oxenius B, Horgan K, Lines C, Riddell R, Morton D, Lanas A, Konstam MA, Baron JA. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. N Engl J Med 2005; 352:1092–1102.
20. Bertagnolli MM, Eagle CJ, Zauber AG, Redston M, Solomon SD, Kim K, Tang J, Rosenstein RB, Wittes J, Corle D, Hess TM, Woloj GM, Boisserie F, Anderson WF, Viner JL, Bagheri D, Burn J, Chung DC, Dewar T, Foley TR, Hoffman N, Macrae F, Pruitt RE, Saltzman JR, Salzberg B, Sylwestrowicz T, Gordon GB, Hawk ET; APC Study Investigators. Celecoxib for the prevention of sporadic colorectal adenomas. N Engl J Med 2006;355:873–884.
21. Key Health Statistics from General Practice 1998: series MB6 (no 2). London, United Kingdom: Office for National Statistics, 2000.
22. Hippisley-Cox J, Pringle M. Prevalence, care and outcomes for patients with diet controlled diabetes in general practice: cross sectional survey. Lancet 2004;364:423–428.
23. Hammersley V, Hippisley-Cox J, Wilson A, Pringle M. A comparison of research general practices and their patients with other practices–cross sectional survey in Trent. Br J Gen Pract 2002; 52:463–468.
24. Hippisley-Cox J, Hammersley V, Pringle M, Coupland C, Crown N, Wright L. How useful are general practice databases for research? Analysis of their accuracy and completeness in one research network. Health Informatics J 2004;10:91–109.
25. Hippisley-Cox J, Coupland C, Logan R. Risk of adverse gastrointestinal outcomes in patients taking cyclo-oxgenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis. BMJ 2005;331:1310–1316.
26. Hippisley-Cox J, Coupland C. Risk of myocardial infarction in patients taking cyclo-oxygenase-2 inhibitors or conventional nonsteroidal anti-inflammatory drugs: population based nested case-control analysis. BMJ 2005;330:1366–1374.
27. Hippisley-Cox J, Coupland C. Effect of combinations of drugs on all cause mortality in patients with ischaemic heart disease: nested case-control analysis. BMJ 2005;330:1059–1063.
28. Jick H, Jick S, Derby LE, Vasilakis C, Myers MW, Meier CR. Calcium-channel blockers and risk of cancer. Lancet 1997;349: 525–528.
29. Eachus J, Williams M, Chan P, Smith GD, Grainge M, Donovan J, Frankel S. Deprivation and cause specific morbidity: evidence from the Somerset and Avon survey of health. BMJ 1996;312: 287–292.
30. Cancer Statistics Registration 2003: series MB1 (no 34). London, United Kingdom: Office for National Statistics, 2005.
31. Chan A, Giovannucci E, Meyerhardt J, Schernhammer E, Curhan G, Fuchs C. Long term use of aspirin and nonsteroidal anti-inflammatory drugs and risk of colorectal cancer. JAMA 2005; 294:914–923.
32. Arber N, Eagle CJ, Spicak J, Rácz I, Dite P, Hajer J, Zavoral M, Lechuga MJ, Gerletti P, Tang J, Rosenstein RB, Macdonald K, Bhadra P, Fowler R, Wittes J, Zauber AG, Solomon SD, Levin B; PreSAP Trial Investigators. Celecoxib for the prevention of colorectal adenomatous polyps. N Engl J Med 2006;355:885–895.
33. Graaf MR, Beiderbeck AB, Egberts ACG, Richel DJ, Guchelaar H-J. The risk of cancer in users of statins. J Clin Oncol 2004;22: 2388–2394.
34. Coogan PF, Smith J, Rosenberg L. Statin use and risk of colorectal cancer. J Natl Cancer Inst 2007;99:32–40.
35. Nazareth I, King M, Haines A, Ranger L, Myers S. Accuracy of diagnosis of psychosis in general practice computer system. BMJ 1993;307:32–34.
36. Pollock A, Vickers N. Reducing DCO registrations through electronic matching of cancer registry data and routine hospital data. Br J Cancer 2000;82:714–717.

Received November 17, 2006. Accepted April 26, 2007.

Address requests for reprints to: Yana Vinogradova, MD, University Park, 13th Floor Tower Building, Nottingham NG7 2RD, United

Kingdom. e-mail: yana.vinogradova@nottingham.ac.uk; fax: (44) 11584 66904.

Supported by the University of Nottingham. This work was undertaken during the course of normal academic duties. The QResearch database undertakes other work for a variety of government organisations. This work was not commissioned or funded by any external organisation.

We acknowledge that the Egton Medical Information Services (EMIS) practices which contribute data to the QRESEARCH database are free of charge. We also acknowledge the contribution of Egton Medical Information Services and David Stables (medical director of EMIS) for support in creating and maintaining the research database.

The study was initiated by JHC and designed by all the authors. The analysis was undertaken by YV and checked by JHC and CC. All authors contributed to the detailed analysis plan, interpretation of results, and drafting of the paper. YV and JHC are the guarantors of the study.

All authors declare that they have no conflict of interest to disclose.

Paper 2

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to statins and risk of common cancers: a series of nested case-control studies. *BMC Cancer*. 2011;11:409.

Vinogradova et al. BMC Cancer 2011, 11:409
http://www.biomedcentral.com/1471-2407/11/409

BMC Cancer

RESEARCH ARTICLE Open Access

Exposure to statins and risk of common cancers: a series of nested case-control studies

Yana Vinogradova*, Carol Coupland and Julia Hippisley-Cox

1 Abstract

Background: Many studies and meta-analyses have investigated the effects of statins on cancer incidence but without showing consistent effects.

Methods: A series of nested case-control studies was conducted covering 574 UK general practices within the QResearch database. Cases were patients with primary cancers diagnosed between 1998 and 2008. The associations between statin use and risk of ten site-specific cancers were estimated with conditional logistic regression adjusted for co-morbidities, smoking status, socio-economic status, and use of non-steroidal anti-inflammatory drugs, cyclo-oxygenase-2 inhibitors and aspirin.

Results: 88125 cases and 362254 matched controls were analysed. The adjusted odds ratio for any statin use and cancer at any site were 1.01 (95%CI 0.99 to 1.04). For haematological malignancies there was a significant reduced risk associated with any statin use (odds ratio 0.78, 95%CI 0.71 to 0.86). Prolonged (more than 4 years) use of statins was associated with a significantly increased risk of colorectal cancer (odds ratio 1.23, 95%CI 1.10 to 1.38), bladder cancer (odds ratio 1.29, 95%CI 1.08 to 1.54) and lung cancer (odds ratio 1.18, 95%CI 1.05 to 1.34). There were no significant associations with any other cancers.

Conclusion: In this large population-based case-control study, prolonged use of statins was not associated with an increased risk of cancer at any of the most common sites except for colorectal cancer, bladder cancer and lung cancer, while there was a reduced risk of haematological malignancies.

2 Background

Multiple randomised controlled trials have demonstrated the benefits of statins in improving survival for patients with ischaemic heart disease [1-5] and this has caused a substantial increase in statin use. While there are definite benefits from statins in reduction of mortality in high risk patients, uncertainties remain about whether statins might increase or decrease the risk of cancer[6-8]. This is important because statins are prescribed for extended periods to large numbers of patients.

The effect of long-term statin use is quite complex because the multiple properties of statins go beyond lipid lowering. There is evidence that statins increase endothelial dysfunction [9] and lower inflammatory markers[10] but it is still not clear whether they may affect the risk of cancer. Experimental data (primarily using rats) have shown both carcinogenicity of statins [11] and no effect on carcinogenesis[12]. Some studies performed on human cancer cells *in vitro* have suggested that statins may be chemo-prophylactic against various types of cancer including colon[13] and breast cancer[14,15]. It has also been found that statins may suppress the growth of cancer cells in vitro by causing the cells to pause in the G1 phase of the mitotic cycle [16] and by increasing cell death[17].

There have been many randomised controlled trials of statins, but cancer has never been a primary outcome. The numbers of cancer cases have been relatively small and the duration of the trials too short to detect the effect of statins on cancer risk. The results from 35 randomised control trials have been summarised in a meta-analysis[18] reporting no association between statin use and overall cancer risk. However, the latest published results of another randomised controlled trial, not included in the meta-analysis, on the use of a combination of simvastatin and ezetimibe in patients with aortic

* Correspondence: yana.vinogradova@nottingham.ac.uk
Division of Primary Care, 13th floor, Tower Building, University Park, Nottingham, NG2 7RD, UK

stenosis demonstrated an increased risk for any cancer (105 vs.70, P = 0.01)[19].

A number of observational studies were designed to assess risk of particular cancers in statin users and the results have been aggregated in a meta-analysis[7]. However, only some of the studies reported statin use of more than 5 years[20]. None of those findings were statistically significant except for one study reporting a decreased risk of prostate cancer, but based only on 42 statin users[21]. A recent study of statin use and ten common cancers[22] found a significantly reduced risk of haematological malignancies and an increased risk of endometrial cancer associated with more than 5 years of statin use.

All studies were smaller than the proposed one, and they were too dissimilar in their definitions of statin use to be analysed together: they either studied different types of statin or statin types were not specified. They also had differing lengths of intervention or follow-up, and included different confounding factors in their analyses.

Given the uncertainty regarding risk of cancer in association with statin usage, we designed a study to determine the risk for the most common incident cancers associated with taking statins including for prolonged periods using a very large population-based research database QRESEARCH. The size of the study has enabled us to adjust for use of other drugs and many potential confounding factors.

3 Methods

3.1 Study design, data source and population

We conducted a series of nested case control studies within a cohort of patients registered with practices in the UK contributing to the QRESEARCH database (version 20). The QResearch database (http://www.qresearch.org) is one of the largest general practice databases containing anonymised clinical records for over 11 million patients registered with 574 UK general practices. The information recorded on the database includes patient demographics (year of birth, sex, socio-demographic data derived from UK census 2001), characteristics (height, weight, smoking status), clinical diagnoses, symptoms, and prescribed medications including repeat prescriptions. The database has been validated by comparing birth rates, death rates, consultation rates, prevalence and mortality rates with other data sources, including the General Household Survey and the General Practice Research Database, and has demonstrated good levels of completeness and consistency [23,24]. Practices were included in the analysis only if they had complete data transmission until at least 1st July 2008.

We identified an open cohort of patients registered with the study practices during the 10 year study period between 1st Jan 1998 and 1st July 2008. We then used READ codes to select all cases aged between 30 and 100 years with a first record of any cancer in the patients' electronic records occurring during the study period. Each case was linked to 5 controls alive and registered with the practice at the time of diagnosis of the case and matched by age, sex, practice and calendar time. Controls were allocated an index date which was the date on which their matched case was first diagnosed with cancer.

3.2 Exclusions

Cases with secondary cancers (READ codes: B56, B57, B58) were excluded. Cases and controls with a diagnosis of any cancer before the index date were excluded. In addition, for breast cancer, we excluded cases and controls with any prior record of mastectomy or prescriptions for tamoxifen since they could be breast cancer cases without a recorded diagnosis in their record. To ensure completeness of exposure data we also excluded temporary residents and patients with fewer than 6 years of medical records before the index date for the main analysis - and fewer than 10 years for the further analysis.

3.3 Primary outcomes

We determined the risks for the most common cancers in the UK[25], comparing these for patients prescribed statins against those not prescribed the drugs. The investigated cancers and corresponding READ codes were: Breast cancer (women, B34), Prostate cancer (men, B46), Lung cancer (B22), Bladder cancer (B49), Haematological malignancies (B6), Gastric cancer (B11), Oesophageal cancer (B10), Colorectal cancer (B13, B14), Pancreatic cancer (B17) and Melanoma (B32). As haematological malignancies cover a range of diseases, possibly differentially affected by statins, we also investigated leukaemia (B63-B6z), lymphoma (B60-B62) and myeloma (B63) separately.

3.4 Exposure variables

Statin exposure was determined based on all prescriptions for statins until 1 year before the index date (date of diagnosis or equivalent date for controls). The drugs included were atorvastatin, pravastatin, fluvastatin, cerivastatin, rosuvastatin, and simvastatin. Prescriptions in the year before the index date were ignored because including these could lead to results being affected by reverse causality - prescribing in cases in this period might be the result of consultations relating to early cancer symptoms before the recorded diagnosis and this could attenuate any protective effect or exaggerate any harmful effect.

For the main analysis, we considered a 60-month period comprising statin prescriptions for the 13 to 72

months prior to the index date. For the additional analysis, covering a follow-up of 10 years, the period considered was 98 months - for the 13 to 120 months prior to the index date.

Statin use was categorised in a number of ways. We considered a patient as a statin user if they had at least 2 prescriptions in the 60-month period (or the 98-month period for the 10-year analysis). We estimated the cumulative use of statins by extracting the duration of use for every prescription and, for groups of prescriptions with inter-prescription gaps of less than 60 days, we calculated overall course times from the start of the first prescription to the end of the last prescription. We then calculated cumulative use as the sum of all overall course times and for the main analysis categorized cumulative use for each patient as: no use; less than 12 months; 13 to 24 months; 25 to 36 months; 37 to 48 months; 49 to 60 months. A test for trend was performed using the actual number of months of use. For the further analysis, covering a follow-up period of 10 years, the categorisation of the time period for statin use was: no use; less than 12 months, 13 to 24 months, 25 to 48 months, 49 to 72 months, and more than 73 months.

If there were at least 2 prescriptions in the 60-month main study period (or in the 98-month additional study period), we conducted analyses for the following individual statin types: simvastatin, atorvastatin, cerivastatin, fluvastatin, pravastatin and rosuvastatin. For the most common types - simvastatin, atorvastatin and pravastatin - we also examined the effect of cumulative use on cancer risk.

Statin dosage was calculated as median dose across the observation time period, and was categorised as low, medium or high according to statin efficacy[26]. The effect of stopping statin usage on risk of cancer was investigated in the main study only by comparing the last prescription date in the study period with the date one year before the index date and categorising as: no statin use in the 13 to 72 months prior to the index date; still on statins; stopped statins 13 to 24 months before the index date; and stopped statins 25 or more months prior to the index date.

3.5 Potential confounding variables

We adjusted for variables which are established cancer risk factors: diabetes[27], rheumatoid arthritis[28], hypertension[29] and body mass index (< 25, 25-29.99, ≥ 30 kg/m^2)[30], if recorded at least 1 year before the index date, and for smoking status (non-smoker, ex-smoker, current smoker) and individual Townsend deprivation score (measure of socio-economic status, in fifths), if recorded before the index date. The Townsend score was based on 2001 postcode-related census data, with higher scores indicating greater level of material deprivation and was used because there is a link between deprivation and incidence of some types of cancer[31]. We adjusted for cardiovascular disease as the main reason for statin therapy. For breast cancer we also accounted for any previous benign breast disease (fibrocystic disease, intraductal papilloma, fibroadenoma) and for family history of breast cancer. For colorectal cancer, additional confounders considered were colitis and Crohn's disease.

We also adjusted for use of traditional non steroidal anti-inflammatory drugs, cyclooxygenase-2 inhibitors and aspirin, as several studies have found protective effects for non-steroidal anti-inflammatory drugs and aspirin on various types of cancer[32,33], in particular on colorectal cancer[33]. We categorised the number of prescriptions for these drugs in the 60-month main study period as: none; 1 to 12; 13 to 24; and 25 or more (adding 25 to 48 and 49 or more for the 98-month additional study period); and adjusted for those categories in assessing cancer risk. We also included in the analyses use of other medications likely to increase the risk of cancer (hormone replacement therapy and oral contraceptive use for breast cancer analysis[34]) if there were at least 2 prescriptions of a drug in the 60-month main study period or 98-month additional study period.

3.6 Statistical analysis

We used conditional logistic regression to estimate odds ratios with 95% confidence intervals for cancer overall and each of the specific cancer sites. As body mass index, smoking status and Townsend deprivation score may be important confounders and have a certain amount of missing data, we used multiple imputation for the missing values[35,36]. We used the ICE procedure in STATA to obtain 5 imputed datasets and applied Rubin's rules to combine effect estimates and estimate standard errors to allow for uncertainty caused by the missing data. We repeated the imputation procedure for each type of cancer separately.

The initial analysis model determined the unadjusted odds ratios for each cancer associated with statin prescriptions according to: any use of statins in the 60-month study period (at least 2 prescriptions in the 13 to 72 months before the index date); cumulative duration of use; and the median prescribed dose. A multivariate model determined the odds ratio for each cancer associated with statin prescriptions adjusted for the potential confounding effects of variables listed above. For comparison with the analyses using imputed data for smoking status and body mass index, we also ran complete case analyses including only cases and controls with complete data as well as analyses using indicator variables for missing categories of smoking, deprivation and body mass index.

We used all the available data on the QResearch database so did not do a pre-study sample size calculation. According to post-hoc calculation, in order to detect an odds ratio of 0.8 (or 1.2) with 80% power at 1% significance for an exposure that occurs in 15% of controls a sample of 2685 cases (or 3424 cases) would be needed. We checked that we had sufficient power for analysis of the six commoner cancers. STATA v 10 was used for all the analyses. We used a 1% significance level to account for the multiple outcomes.

4 Results

Overall there were 118,780 patients with a recorded diagnosis of cancer at any site within the study period. 3,810 patients had diagnoses of secondary cancers so were removed from the analysis. Thirty six patients were coded with cancers applicable only to the other gender and were also removed. For breast cancer 370 cases and 302 controls with a previous history of mastectomy were excluded as were a further 685 cases and 471 controls with a previous history of tamoxifen use. This left a total of 113,879 patients with a first diagnosis of cancer during the study period and 568,958 controls. After removing patients with less than 6 years of medical records there were 88,125 cases of primary cancer matched with 362,254 controls. Eighty-one percent of cases and 71% of controls also had complete data for 10 years of follow-up. The proportions of cases with different types of cancer were similar to proportions in cancer registration statistics in England for 2003[37] for patients older than 30 years.

4.1 Baseline characteristics

Table 1 shows baseline characteristics for cases of cancer at any site and their matched controls. Fifty-three percent of the cases were men; the median age at diagnosis was 69 years (interquartile range: 60 to 77). Seventy six percent of cases and 73% of controls had complete data for body mass index, smoking status and Townsend deprivation score.

Cases and controls had similar patterns of co-morbidity except for diabetes (8.1% in cases vs. 7.4% in controls). The difference in proportion of diabetic patients was most marked in pancreatic cancer cases (12.7% vs. 8.3% in controls).

4.2 Statin exposure

Overall 15.5% of cases and 15.1% of controls had at least 2 statin prescriptions between 13 to 72 months prior to the index date. Most of the statin users (95% of cases and controls) had statin prescriptions for more than a year. Median numbers of scripts for statin users were 19 (interquartile ranges, 9 to 32) for cases and for controls. Median numbers of months on statin were 28 for cases and controls (interquartile ranges, 12 to 50 for cases and 12 to 49 for controls).

The most frequently prescribed statins were simvastatin (9.2% of cases and 9.0% of controls), atorvastatin (6.1% of cases and 5.9% of controls) and pravastatin (1.6% of cases and controls). The other statins were prescribed to less than 1% of the population. Very few atorvastatin users had low dose prescriptions (3 cases and 28 controls) and few pravastatin users had high dose prescriptions (4 cases and 12 controls). Simvastatin dosage was distributed evenly. Long-term statin use was associated with higher dose: in patients prescribed statins for more than 4 years, 42% of cases and 43% controls were on high doses compared with 31% cases and 31% controls on high doses in patients prescribed statins for less than 4 years.

The results of the main analyses, based on patients with at least 6 years of medical records, are shown in Tables 2, 3, 4 and Figure 1. Table 5 shows the odds ratios for each cancer according to cumulative duration of statin use in patients with at least 10 years of medical records.

4.2.1 Cancer of any site

The analysis for overall risk of cancer (at any site) did not show a significant association with any statin use (Table 2). Patients with a cumulative prescription duration of more than one year had a similar risk of cancer of any site compared with patients with no statin prescriptions (adjusted odds ratio (AOR) 1.02, 95%CI 0.99 to 1.05). Analyses of trends for duration of use (Table 3) and dosage, as well as analysis of use of individual statins (Table 4), did not show any effect of statins on overall risk of cancer.

4.2.2 Colorectal cancer

There was no overall increase of colorectal cancer risk in statin users (AOR 1.07, 95%CI 1.00 to 1.15, P = 0.056), with a slight association for patients with prescriptions for more than a year (AOR = 1.09, 95%CI 1.01 to 1.18, P = 0.036), which was not however statistically significant at P < 0.01. Further analysis showed a significant association with duration of use of statins (P_{trend}= 0.001), with a 23% increased risk for 49 to 60 months of use of (AOR 1.23, 95%CI 1.10 to 1.38) compared with no use. The analysis of the median prescribed dose of statins revealed a significant association with an 18% increased risk on high dose of statin (AOR 1.18, 95%CI 1.07 to 1.31, P = 0.001).

Analyses of individual statins showed an association between colorectal cancer and atorvastatin (P_{trend}= 0.001), with an increased risk of colorectal cancer associated with atorvastatin use of 4 or more years (AOR 1.51, 95%CI 1.24 to 1.83).

The risk of colorectal cancer was not significantly increased for patients who stopped taking statins more

Table 1 Baseline characteristics for all cases with primary cancer and their matched controls with at least 6 years of medical records

	Cases (N = 88125)	Controls (N = 362254)
Sex		
female	41749 (47.4)	170173 (47.0)
male	46376 (52.6)	192081 (53.0)
Age band (years)		
30-54	13151 (14.9)	49906 (13.8)
55-64	19638 (22.3)	80107 (22.1)
65-74	26758 (30.4)	111698 (30.8)
75-84	25013 (28.4)	106278 (29.3)
85 +	3565 (4.0)	14265 (3.9)
Deprivation		
Townsend quintile 1, most affluent	22072 (25.0)	92287 (25.5)
Townsend quintile 2	18998 (21.6)	79067 (21.8)
Townsend quintile 3	17338 (19.7)	71358 (19.7)
Townsend quintile 4	15325 (17.4)	61767 (17.1)
Townsend quintile 5, most deprived	11896 (13.5)	45971 (12.7)
Townsend missing	2496 (2.8)	11804 (3.3)
Body mass index (kg/m^2)		
15-24	26721 (30.3)	105883 (29.2)
25-29	27285 (31.0)	108803 (30.0)
30-49	12922 (14.7)	51413 (14.2)
not recorded	21197 (24.1)	96155 (26.5)
Smoking status		
non-smoker	54307 (61.6)	233135 (64.4)
ex-smoker	7567 (8.6)	23842 (6.6)
current smoker	17275 (19.6)	54869 (15.1)
not recorded	8976 (10.2)	50408 (13.9)
Co-morbidities		
Cardiovascular disease	14278 (16.2)	58123 (16.0)
Diabetes	7115 (8.1)	26802 (7.4)
Hypertension	27104 (30.8)	109797 (30.3)
Osteoarthritis	12807 (14.5)	52586 (14.5)
Rheumatoid arthritis	1310 (1.5)	5132 (1.4)
Colitis[1]	124 (1.1)	293 (0.6)
Crohn's disease[1]	28 (0.2)	109 (0.2)
Benign breast disease[2]	1094 (7.0)	2937 (4.7)
Family history of breast cancer[2]	539 (3.4)	1249 (2.0)
Medications (in previous 13-72 months)		
NSAID	35697 (40.5)	140642 (38.8)
COX2 inhibitors	6901 (7.8)	26974 (7.4)
Aspirin	19895 (22.6)	79067 (21.8)
Hormone replace therapy[2]	3289 (21.0)	10973 (17.4)
Oral contraceptive pill[2]	523 (3.3)	1638 (2.6)

[1)] Based only on cases with colorectal cancer and their controls

[2)] Based only on female cases with breast cancer and their controls

Table 2 Use of statins in cases and controls in 13 to 72 months prior the index date by cancer site (in cases and matched controls with at least 6 years of medical records)

Cancer	Total number of cases	Total number of controls	N of statin users in cases (%)	N of statin users in controls (%)	Unadjusted OR (95% CI)		Adjusted[#] OR (95% CI)		P-value
breast†	15666	62938	1481 (9.5)	6227 (9.9)	0.98	(0.92 to 1.04)	1.00	(0.93 to 1.08)	0.993
prostate	14764	61853	2774 (18.8)	11508 (18.6)	1.03	(0.98 to 1.08)	1.08	(1.01 to 1.14)	0.016
colorectal‡	11749	48624	2000 (17.0)	7770 (16.0)	1.12	(1.06 to 1.19)	1.07	(1.00 to 1.15)	0.056
lung	10163	42415	1998 (19.7)	7621 (18.0)	1.16	(1.09 to 1.23)	1.07	(0.99 to 1.16)	0.095
blood	7185	29162	973 (13.5)	4339 (14.9)	0.91	(0.84 to 0.99)	0.78	(0.71 to 0.86)	< 0.001
bladder	4227	17559	856 (20.3)	3125 (17.8)	1.23	(1.12 to 1.34)	1.15	(1.03 to 1.29)	0.012
skin	3249	13115	433 (13.3)	1675 (12.8)	1.12	(0.99 to 1.26)	1.08	(0.93 to 1.26)	0.292
oesophagus	3159	13041	496 (15.7)	2106 (16.1)	0.97	(0.87 to 1.09)	0.88	(0.77 to 1.01)	0.072
pancreas	2110	8762	365 (17.3)	1397 (15.9)	1.15	(1.01 to 1.32)	0.96	(0.82 to 1.14)	0.671
stomach	1992	8279	322 (16.2)	1363 (16.5)	1.00	(0.87 to 1.16)	0.86	(0.72 to 1.02)	0.078
All cancers	**88125**	**362254**	**13621 (15.5)**	**54606 (15.1)**	**1.07**	**(1.05 to 1.09)**	**1.01**	**(0.99 to 1.04)**	**0.280**

[#] Adjusted for Townsend quintile, body mass index, smoking status, myocardial infarction, coronary heart disease, diabetes, hypertension, stroke, rheumatoid arthritis, use of NSAIDs, Cox2-inhibitors, aspirin

† Also adjusted for family history of breast cancer, use of oral contraceptives, hormone-replace therapy

‡ Also adjusted for colitis and Crohn's disease

than 2 years before the index date (AOR 1.03, 95%CI 0.81 to 1.31).

The increased risk of colorectal cancer associated with longer duration of statin use found in patients with at least 6 years of medical records was not supported by the trend test of months on medication in patients with at least 10 years of medical records (p = 0.069).

4.2.3 Bladder cancer

For bladder cancer, there was a borderline 15% increased risk of cancer associated with any use of statins (P = 0.012) and a 16% increased risk associated with more than one year's use (P = 0.018), but these were not statistically significant. For patients with more than 48 months of statin use, risk of bladder cancer was 29% higher (AOR, 1.29, 95% 1.08 to 1.54, P = 0.006) but the trend test for duration was not statistically significant (P_{trend} = 0.014). No particular type of statin was significantly associated with an increased risk. The risk of bladder cancer was not significantly increased in patients who stopped taking statins more

Table 3 Cumulative duration of statin use in cases and controls in 13 to 72 months prior to the index date by cancer site (in cases and matched controls with at least 6 years of medical records)

	Less than 12 months		13 to 24 months		25 to 48 months		49 months and more		
cancer	**Cases/ Controls**	**Adjusted Odds ratio (95%CI) [#]**	**Cases/ Controls**	**Adjusted Odds ratio (95%CI) [#]**	**Cases/ Controls**	**Adjusted Odds ratio (95%CI) [#]**	**Cases/ Controls**	**Adjusted Odds ratio (95%CI) [#]**	**P-* value**
breast†	433/1811	1.01 (0.90 to 1.13)	289/1292	0.93 (0.81 to 1.07)	430/1685	1.07 (0.95 to 1.21)	329/1439	0.95 (0.83 to 1.09)	0.719
prostate	668/2784	1.05 (0.95 to 1.15)	560/2187	1.14 (1.03 to 1.26)	796/3295	1.09 (0.99 to 1.19)	750/3242	1.05 (0.95 to 1.16)	0.084
colorectal‡	525/2038	1.05 (0.95 to 1.17)	400/1595	1.04 (0.92 to 1.17)	539/2230	1.02 (0.92 to 1.14)	536/1907	1.23 (1.10 to 1.38)	0.002
lung	485/1857	1.02 (0.90 to 1.15)	406/1478	1.11 (0.97 to 1.27)	549/2233	1.01 (0.90 to 1.14)	558/2053	1.18 (1.05 to 1.34)	0.013
blood	255/1082	0.84 (0.72 to 0.98)	201/860	0.81 (0.68 to 0.96)	277/1307	0.73 (0.63 to 0.84)	240/1090	0.76 (0.65 to 0.89)	< 0.001
bladder	209/785	1.13 (0.95 to 1.34)	174/611	1.18 (0.98 to 1.42)	240/952	1.06 (0.90 to 1.25)	233/777	1.29 (1.08 to 1.54)	0.014
skin	120/422	1.19 (0.95 to 1.49)	61/347	0.74 (0.55 to 0.99)	141/474	1.23 (0.99 to 1.54)	111/432	1.08 (0.84 to 1.39)	0.373
oesophagus	126/571	0.82 (0.67 to 1.02)	97/394	0.91 (0.71 to 1.17)	128/601	0.82 (0.66 to 1.02)	145/540	1.04 (0.83 to 1.30)	0.705
pancreas	87/367	0.85 (0.66 to 1.10)	73/269	1.02 (0.76 to 1.36)	113/390	1.09 (0.86 to 1.40)	92/371	0.97 (0.74 to 1.28)	0.521
stomach	76/317	0.87 (0.66 to 1.15)	69/271	0.90 (0.67 to 1.20)	94/404	0.86 (0.67 to 1.11)	83/371	0.85 (0.64 to 1.12)	0.167
All cancers	**3467/ 13935**	**1.01 (0.97 to 1.05)**	**2752/ 10855**	**1.02 (0.98 to 1.07)**	**3868/ 15708**	**1.00 (0.96 to 1.04)**	**3534/ 14108**	**1.04 (1.00 to 1.09)**	**0.057**

[#] Adjusted for Townsend quintile, body mass index, smoking status, myocardial infarction, coronary heart disease, diabetes, hypertension, stroke, rheumatoid arthritis, use of NSAIDs, Cox2-inhibitors, aspirin

† Also adjusted for family history of breast cancer, benign breast disease, use of oral contraceptives, hormone-replace therapy

‡ Also adjusted for colitis and Crohn's disease

* Trend test based on number of months prescribed

Table 4 Types of statins in cases and controls in 13 to 72 months prior to the index date (in cases and matched controls with at least 6 years of medical records)

	Atorvastatin			Pravastatin			Simvastatin		
cancer	**Cases/ Controls**	**Adjusted Odds ratio (95%CI) [#]**	**P-value**	**Cases/ Controls**	**Adjusted Odds ratio (95%CI) [#]**	**P-value**	**Cases/ Controls**	**Adjusted Odds ratio (95%CI) [#]**	**P-value**
breast†	596/2574	0.96 (0.86 to 1.06)	0.387	152/630	1.02 (0.85 to 1.23)	0.835	871/3720	0.98 (0.89 to 1.07)	0.584
prostate	1023/4398	0.99 (0.91 to 1.07)	0.781	314/1182	1.15 (1.00 to 1.31)	0.046	1668/6924	1.05 (0.99 to 1.13)	0.117
colorectal‡	826/2934	1.17 (1.07 to 1.28)	0.001	212/786	1.09 (0.93 to 1.28)	0.289	1152/4783	0.96 (0.88 to 1.04)	0.273
lung	786/2912	1.07 (0.97 to 1.18)	0.179	195/837	0.93 (0.78 to 1.11)	0.435	1205/4588	1.06 (0.97 to 1.15)	0.202
blood	381/1653	0.87 (0.77 to 0.99)	0.041	95/452	0.84 (0.66 to 1.06)	0.138	579/2592	0.82 (0.73 to 0.91)	< 0.001
bladder	353/1212	1.19 (1.03 to 1.37)	0.015	87/309	1.08 (0.84 to 1.40)	0.544	513/1893	1.10 (0.97 to 1.25)	0.119
skin	168/655	1.03 (0.84 to 1.25)	0.805	43/178	0.91 (0.64 to 1.30)	0.609	259/998	1.06 (0.90 to 1.26)	0.483
oesophagus	197/846	0.88 (0.73 to 1.05)	0.159	62/226	1.05 (0.78 to 1.42)	0.740	298/1222	0.94 (0.80 to 1.09)	0.403
pancreas	143/557	0.92 (0.74 to 1.14)	0.439	37/152	0.92 (0.62 to 1.35)	0.667	224/852	0.99 (0.83 to 1.20)	0.952
stomach	123/500	0.94 (0.75 to 1.18)	0.604	41/149	1.03 (0.71 to 1.50)	0.880	186/819	0.85 (0.70 to 1.03)	0.106
All cancers	**5357/ 21253**	**1.01 (0.98 to 1.05)**	**0.461**	**1442/ 5680**	**1.02 (0.96 to 1.09)**	**0.488**	**8102/ 32769**	**1.00 (0.97 to 1.03)**	**0.844**

[#] Adjusted for Townsend quintile, body mass index, smoking status, myocardial infarction, coronary heart disease, diabetes, hypertension, stroke, rheumatoid arthritis, use of NSAIDs, Cox2-inhibitors, aspirin

† Also adjusted for family history of breast cancer, benign breast disease, use of oral contraceptives, hormone-replace therapy

‡ Also adjusted for colitis and Crohn's disease

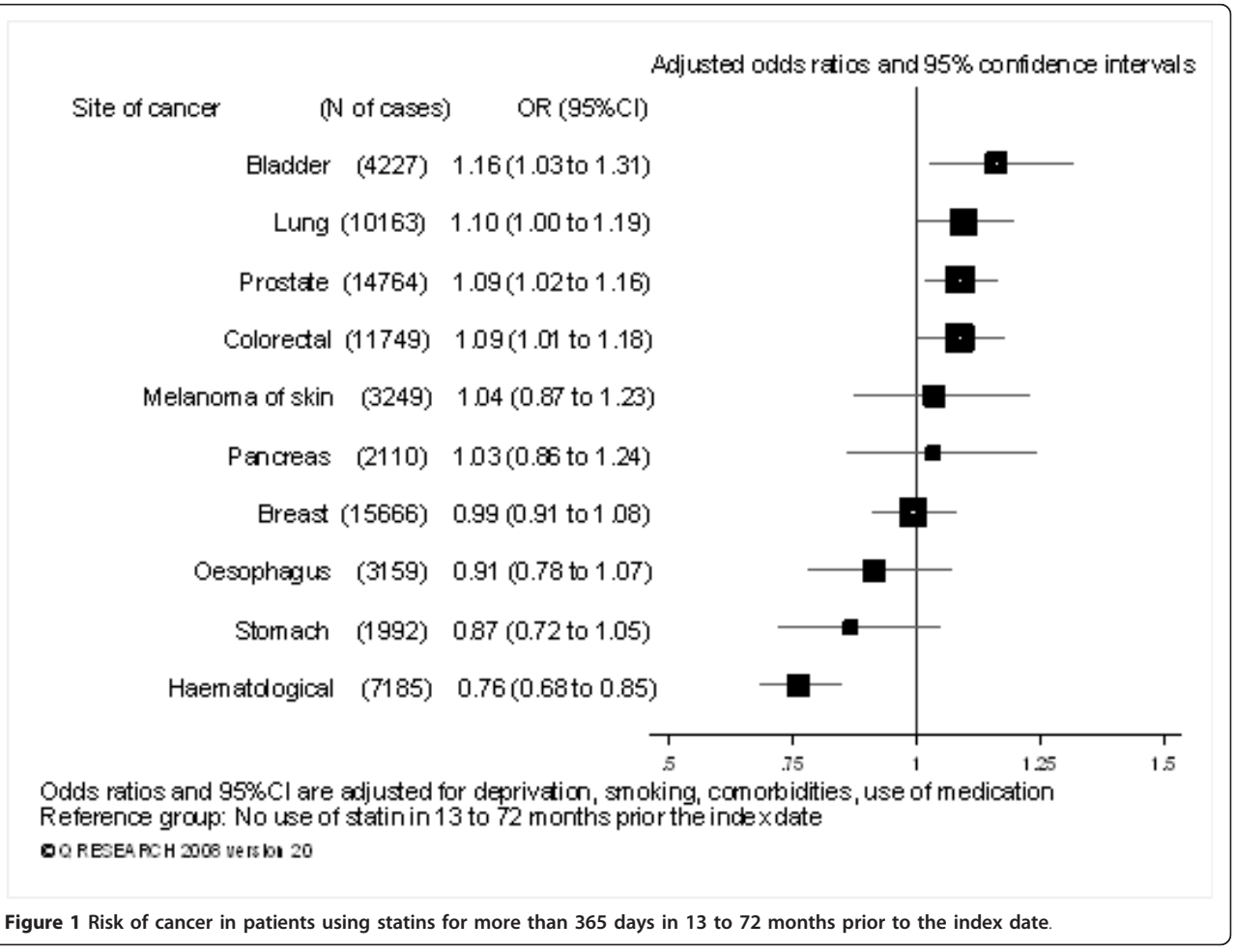

Figure 1 Risk of cancer in patients using statins for more than 365 days in 13 to 72 months prior to the index date.

Table 5 Cumulative duration of statin use in cases and controls in 13 to 120 months prior to the index date by cancer site in cases and controls with 10 or more years of recorded data

	Less than 12 months		13 to 24 months		25 to 48 months		49 to 72 months		73 months and more		
cancer	**Cases/ Controls**	**Adjusted Odds ratio (95%CI) #**	**Cases/ Controls**	**Adjusted Odds ratio (95%CI) #**	**Cases/ Controls**	**Adjusted Odds ratio (95%CI) #**	**Cases/ Controls**	**Adjusted Odds ratio (95%CI) #**	**Cases/ Controls**	**Adjusted Odds ratio (95%CI)#**	**P-* value**
breast†	363/1338	1 (0.86 to 1.16)	254/962	0.96 (0.81 to 1.15)	346/1208	1.06 (0.90 to 1.25)	169/634	0.93 (0.75 to 1.15)	131/542	0.85 (0.67 to 1.08)	0.220
prostate	545/2119	0.99 (0.87 to 1.12)	489/1609	1.19 (1.04 to 1.36)	641/2424	1.06 (0.94 to 1.19)	369/1394	1.07 (0.93 to 1.24)	320/1234	1.12 (0.96 to 1.32)	0.173
colorectal‡	446/1527	1.06 (0.92 to 1.23)	327/1224	1.03 (0.88 to 1.21)	429/1625	0.97 (0.84 to 1.13)	285/902	1.20 (1.01 to 1.43)	213/711	1.21 (0.99 to 1.48)	0.069
lung	419/1375	1.06 (0.91 to 1.25)	342/1064	1.21 (1.01 to 1.43)	458/1586	1.05 (0.89 to 1.23)	259/950	1.00 (0.82 to 1.21)	235/734	1.17 (0.95 to 1.45)	0.240
blood	202/821	0.72 (0.59 to 0.88)	160/616	0.8 (0.64 to 1.00)	233/966	0.74 (0.61 to 0.89)	133/468	0.87 (0.68 to 1.11)	90/428	0.55 (0.41 to 0.73)	< 0.001
bladder	172/565	1.14 (0.91 to 1.43)	137/456	1.19 (0.93 to 1.53)	190/681	1.01 (0.81 to 1.26)	117/360	1.19 (0.91 to 1.57)	94/280	1.37 (1.02 to 1.86)	0.062
skin	110/312	1.41 (1.04 to 1.89)	56/270	0.79 (0.55 to 1.13)	112/349	1.21 (0.89 to 1.64)	50/176	0.87 (0.56 to 1.36)	56/185	1.04 (0.70 to 1.55)	0.626
oesophagus	103/409	0.9 (0.67 to 1.20)	75/272	1.01 (0.72 to 1.42)	105/445	0.81 (0.61 to 1.09)	68/248	1.09 (0.77 to 1.55)	59/185	1.03 (0.70 to 1.52)	0.870
pancreas	76/277	0.89 (0.63 to 1.24)	61/196	1.22 (0.84 to 1.78)	86/286	0.99 (0.71 to 1.38)	50/172	1.04 (0.69 to 1.56)	40/148	0.85 (0.53 to 1.37)	0.475
stomach	58/232	0.84 (0.58 to 1.22)	60/206	0.82 (0.56 to 1.19)	77/301	0.86 (0.61 to 1.21)	35/150	0.56 (0.35 to 0.90)	38/165	0.63 (0.39 to 1.00)	0.008
All cancers	**2890/ 10413**	**0.99 (0.94 to 1.05)**	**2303/ 8005**	**1.05 (0.99 to 1.11)**	**3107/ 11437**	**0.98 (0.93 to 1.03)**	**1776/ 6319**	**1.01 (0.95 to 1.08)**	**1462/ 5259**	**1.02 (0.95 to 1.10)**	**0.958**

Adjusted for Townsend quintile, body mass index, smoking status, myocardial infarction, coronary heart disease, diabetes, hypertension, stroke, rheumatoid arthritis, use of NSAIDs, Cox2-inhibitors, aspirin

† Also adjusted for family history of breast cancer, benign breast disease, use of oral contraceptives, hormone-replace therapy

‡ Also adjusted for colitis and Crohn's disease

* Trend test based on number of months prescribed

than 2 years before the index date (AOR 0.94, 95%CI 0.62 to 1.40).

The additional analysis restricted to patients with at least 10 years of medical records showed similar results, but these were not statistically significant.

4.2.4 Lung cancer

Although the unadjusted risk of lung cancer appeared to be significantly higher in statin users (unadjusted odds ratio (UOR) 1.16, 95%CI 1.09 to 1.23, P < 0.001), after adjusting for cardiovascular disease the association became much weaker (OR 1.07, 95%CI 1.00 to 1.14, P = 0.067) and did not noticeably change after further adjusting for other factors.

The unadjusted trend test for months of statin use was significant (P < 0.001) and use of statins for more than 4 years was associated with an increased risk of cancer (UOR 1.22, 95%CI 1.10 to 1.35, P < 0.001). After adjusting for cardiovascular disease and other factors, these associations were also reduced but long-term usage remained significant (AOR 1.18, 95%CI 1.05 to 1.34, P = 0.007).

Analyses repeated on patients with 10 years of medical records did not show any statistically significant effect of statins for either overall or long term use.

4.2.5 Prostate cancer

Although the analysis demonstrated an 8% increased risk of prostate cancer for overall statin user and a 9% increased risk for patients with prescriptions covering more than a year, these associations were not statistically significant (P = 0.016 and P = 0.011). There were no dose or duration relationships in patients with either 6 years or 10 years of medical records.

4.2.6 Haematological malignancies

There was a 22% reduced blood cancer risk for overall statin use (AOR 0.78, 95%CI 0.71 to 0.86, P < 0.001) and a 24% reduction for patients with statin prescriptions of more than a year (AOR 0.76, 95%CI 0.68 to 0.85), with a significant trend for duration of use (P_{trend} < .001). No differential effects were found for particular types of statin. Patients who stopped taking statins for more than 2 years had the same risk of cancer as non-statin users (AOR 0.90, 95%CI 0.67 to 1.23).

Although lymphoma, myeloma and leukaemia were similarly associated with overall use of statins and use for more an year, only leukaemia had associations with duration and dose with significant trend tests (P_{trend} = 0.002 and P_{trend} < 0.001), a 26% risk reduction (AOR 0.74, 95%CI 0.62 to 0.87, P = 0.001) with at least two years of statin prescriptions, and a 25% risk reduction on high dose (AOR 0.75, 95%CI 0.61 to 0.92, P = 0.006).

4.2.7 Other cancers

There were no significant associations with statin use for any other cancers.

4.2.8 Sensitivity analyses

The sensitivity analyses treating missing values for smoking, and body mass index as separate categories produced very similar results. The complete case analyses resulted in very similar odds ratios, but the confidence intervals were wider due to the reduced number of observations (results available from the authors).

5 Discussion

In this large population-based case control study to determine the risk of common cancers associated with use of statins, we confirmed that use of statins does not affect the overall risk of cancer. We did find some evidence of an increased risk of colorectal cancer in patients using statins for 4 or more years or with a high statin dose. We also found an increased risk of bladder cancer and lung cancer in patients prescribed statins for 4 or more years. Conversely, we found a reduced risk of haematological malignancies in statin users.

There are a large number of studies devoted to statins and cancer risk summarised in meta-analyses[6-8] which did not show an adverse or protective effect of statins on the overall incidence of cancer. However, the categorisation of 'any cancer' is not a specific enough endpoint of study as it covers a range of diseases, each with a different aetiology and course of development.

Colorectal cancer, as one of the most common cancers, has been studied extensively but only eight epidemiological studies looked at the effect of long-term statin use (at least 4 years). Four of them[38-41] had odds ratios greater than unity (from 1.00 to 1.15) and four of them[22,42-44] reported odds ratios less than unity (from 0.71 to 0.83), but none of these findings reached statistically significant levels even at the 5% level. The effect of dose in our study might, however, be a replication of the effect of cumulative use because a high dose was more likely to be prescribed for patients who had been on statins for substantial period of time.

The other two most common cancers, breast and prostate, also account for a number of studies but there has been no definite outcome in associating any of these with use of statins and our null results are consistent with this. Studies for prostate cancer have been aggregated into a meta-analysis [45], which did not find any significant association with overall risk of prostate cancer and another meta-analysis[46] looking at breast cancer studies also failed to demonstrate a protective or adverse effect of statins.

For bladder cancer, results of a meta-analysis considering 5 studies showed an increased, but not significant, association between stain use and cancer risk[7]. There have been very few studies investigating the long-term

effect of statin use on bladder cancer. One study[39] showed an increased risk for more than 5 years of statin use, which is consistent with our findings, but another very recent one found no significant association for current use of statins for more than 5 years [22]. Both studies, however, were much smaller.

Our findings of a significant increase in unadjusted lung cancer risk for statin use and for long-term use were both significantly decreased by adjusting for cardiovascular disease, but after adjusting for all factors, long-term use still showed a significant association with increased lung cancer risk. There is no causal link between cardiovascular disease and lung cancer but there is a strong association of both conditions with smoking. The finding about possible increased risk from long-term use is consistent with the results of two other studies[22,39], although their findings were not significant.

The decreased risk of haematological malignancies could be explained by reverse causality, as patients with such diagnoses are more likely to have lower lipid levels [47] although we did restrict our statin exposure to prescriptions at least 12 months before diagnosis. The effect of statins on leukaemia has been studied *in vitro* and there is evidence that statins might suppress the growth of promyelocitic[48] and lymphocytic[49] leukemic cells. However, no epidemiological studies have provided significant evidence of any statin effect on incidence of leukaemia.

Our study has several strengths. It is substantially larger and has greater statistical power than any previous study. This has allowed us to perform the analyses separately for different cancers within the same population. We had a substantial number of patients with at least 10 years of records, which also allowed us to examine long-term statin use. The study is based on computer-recorded prescribing and morbidity data collected prospectively. The study was not subject to response bias or recall bias as the exposure data were recorded before the date of diagnosis or pseudo-diagnosis. Any bias from misclassification is likely to be small because the level of accuracy and completeness of medical records in general practices has been shown to be high [50].

Matching the controls on sex, age, practice and calendar year removed confounding by these factors. Any bias from misclassification of statin use is likely to be minimal as more than 99% of all general practitioners' repeat prescriptions are recorded on computer[51]. We minimised the possibility of misleading data from the effects of undiagnosed cancer in new medical records by excluding prescriptions, diagnoses of co-morbidities and records of body mass index made within the 12 months prior to the date of the diagnosis or pseudo-diagnosis of cancer.

Our study has some limitations. Information on certain risk factors for cancer, such as level of physical activity, alcohol use, and diet, and information on cancer screening tests (mammography, prostate-specific antigen test and colonoscopy) were not reliably recorded on the database and not included in the analysis so there may be some residual confounding. Although we adjusted the risk of cancer for possible effects of smoking, obesity, deprivation, co-morbidities and the use of other medications, residual confounding may also result from misclassification of those variables. Values of body mass index or smoking status, were missing for about 22% of cases and 25% of controls, so we substituted missing values using multiple imputation. We did not include blood test results in the analysis, in particular high-density lipoproteins and total-serum cholesterol, because they were not consistently recorded on the data base and would be more likely to be recorded in statin users.

Although our data contain detailed information on drug prescriptions, this may not reflect actual use. However there is no reason to think that any non-adherence would systematically differ between cases and controls.

Another possible source of misclassification arises from a statin (simvastatin 10 mg) having become available over the counter in May 2004 in the UK, which would affect mostly younger people who are not entitled to free prescriptions[52] and only a small part of the study period. However, among statin users 81.4% of cases and 82.4% of controls were aged 65 years or older and therefore entitled to free prescribed medications. Analyses repeated on this group of patients obtained similar results, which suggests that any misclassification of use of medication because of over-the-counter purchase is not an explanation for our findings.

6 Conclusion

In summary, we have conducted a large population-based case-control study that examined the effect of statins on the risk of cancer and found that there is no effect from prolonged use of statins on overall risk of cancer, but that prolonged use of statins may be associated with an increased risk of colorectal cancer, bladder cancer and lung cancer and a decreased risk of haematological malignancies.

7 Approvals

This project has been approved by the QRESEARCH scientific board and notified to the Trent Multi Centre Research Ethics Committee.

8 Funding

There was no external funding for the study.

11 Competing interests

JHC is codirector of QResearch which is a not-for-profit partnership between the University of Nottingham and EMIS.

9 Authors' contributions

YV contributed to the study design, data extraction, data manipulation, data analysis, interpretation and drafting of the paper. JHC had the original idea for this study and extracted the data, contributed to the interpretation and drafting of the paper. CC contributed to the development of the idea, interpretation and drafting of the paper. YV is the guarantor of the study. All authors read and approved the final manuscript.

9 Acknowledgements

We acknowledge the contribution of EMIS and the University of Nottingham for expertise in creating and maintaining QRESEARCH and to the EMIS practices which contribute data without whom this research would not be possible.

Received: 5 May 2011 Accepted: 26 September 2011
Published: 26 September 2011

References

1. Scandinavian Simvastatin Survival Group: **Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S).** *Lancet* 1994, **344**:1383-1389.
2. Long-term Intervention with Pravastatin in Ischaemic Heart Disease (LIPID) Study Group: **Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels.** *NEJM* 1998, **339**:1349-1357.
3. Sacks FM, Pfeffer MA, Moye LA: **The effects of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. The CARE study.** *NEJM* 1996, **335**:1001-1009.
4. Heart Protection Study Collaborative Group: **MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high risk individuals: a randomised placebo-controlled trial.** *Lancet* 2002, **360**:7-22.
5. Cannon C, Braunwald E, McCabe C, Rader D, Rouleau J, Belder R, Joyal S, Hill K, Pfeffer MA, Skene A: **Comparison of intensive and moderate lipid lowering with statins after acute coronary syndromes.** *NEJM* 2004, 350.
6. Dale K, Coleman C, Hentyan N, Kluger J, White C: **Statins and cancer risk - a meta-analysis.** *JAMA* 2006, **295**:74-80.
7. Kuoppala J, Lamminpää A, Pukkala E: **Statins and cancer: A systematic review and meta-analysis.** *European journal of cancer* 2008, **44**:2122-2132.
8. Brugts JJ, Yetgin T, Hoeks SE, Gotto AM, Shepherd J, Westendorp RGJ, de Craen AJM, Knopp RH, Nakamura H, Ridker P, *et al*: **The benefits of statins in people without established cardiovascular disease but with cardiovascular risk factors: meta-analysis of randomised controlled trials.** *BMJ* 2009, **338**:b2376.
9. John S, Schneider MP, Delles C, Jacobi J, Schmieder RE: **Lipid-independent effects of statins on endothelial function and bioavailability of nitric oxide in hypercholesterolemic patients.** *American Heart Journal* 2005, **149**:473.e471-473.e410.
10. Endres M: **Statins: Potential new indications in inflammatory conditions.** *Atherosclerosis Supplements* 2006, **7**:31-35.
11. Newman T, Hulley S, Task Force on Risk, Tobert J, Pahor M, Laws A, Illingworth DR, Cattley R: **Carcinogenicity of Lipid-Lowering Drugs.** *JAMA* 1996, **275**:1481-1482.
12. Lubet RA, Boring D, Steele VE, Ruppert JM, Juliana MM, Grubbs CJ: **Lack of Efficacy of the Statins Atorvastatin and Lovastatin in Rodent Mammary Carcinogenesis.** *Cancer Prev Res* 2009, **2**:161-167.
13. Carlberg M, Dricu A, Blegen H, Wang M, Hjertman M, Zickert P, Hoog A, Larsson O: **Mevalonic acid is limiting for N-linked glycosylation and translocation of the insulin-like growth factor-1 receptor to the cell surface. Evidence for a new link between 3-hydroxy-3-methylglutaryl-coenzyme a reductase and cell growth.** *J Biol Chem* 1996, **271**:17453-17462.
14. Addeo R, Altucci L, Battista T, Bonapace IM, Cancemi M, Cicatiello L, Germano D, Pacilio C, Salzano S, Bresciani F, Weisz A: **Stimulation of human breast cancer MCF-7 cells with estrogen prevents cell cycle arrest by HMG-CoA reductase inhibitors.** *Biochemical & Biophysical Research Communications* 1996, **220**:864-870.
15. Inano H, Suzuki K, Onoda M, Wakabayashi K: **Anti-carcinogenic activity of simvastatin during the promotion phase of radiation-induced mammary tumorigenesis of rats.** *Carcinogenesis* 1997, **18**:1723-1727.
16. Keyomarsi K, Sandoval L, Band V, Pardee AB: **Synchronization of tumor and normal cells from G1 to multiple cell cycles by lovastatin.** *Cancer Research* 1991, **51**:3602-3609.
17. Agarwal B, Bhendwal S, Halmos B, Moss SF, Ramey WG, Holt PR: **Lovastatin augments apoptosis induced by chemotherapeutic agents in colon cancer cells.** *Clin Cancer Res* 1999, **5**:2223-2229.
18. Bonovas S, Filioussi K, Tsavaris N, Sitaras NM: **Statins and Cancer Risk: A Literature-Based Meta-Analysis and Meta-Regression Analysis of 35 Randomized Controlled Trials.** *J Clin Oncol* 2006, **24**:4808-4817.
19. Rossebo AB, Pedersen TR, Boman K, Brudi P, Chambers JB, Egstrup K, Gerdts E, Gohlke-Barwolf C, Holme I, Kesaniemi YA, *et al*: **Intensive Lipid Lowering with Simvastatin and Ezetimibe in Aortic Stenosis.** *N Engl J Med* 2008, **359**:1343-1356.
20. Boudreau DM, Yu O, Johnson J: **Statin use and cancer risk: a comprehensive review.** *Expert Opinion on Drug Safety* 2010, **9**:603-621.
21. Flick ED, Habel LA, Chan KA, Van Den Eeden SK, Quinn VP, Haque R, Orav EJ, Seeger JD, Sadler MC, Quesenberry CP, *et al*: **Statin Use and Risk of Prostate Cancer in the California Men's Health Study Cohort.** *Cancer Epidemiology Biomarkers & Prevention* 2007, **16**:2218-2225.
22. Jacobs EJ, Newton CC, Thun MJ, Gapstur SM: **Long-term Use of Cholesterol-Lowering Drugs and Cancer Incidence in a Large United States Cohort.** *Cancer Research* 2011, **71**:1763-1771.
23. National Statistics: **Key Health Statistics from General Practice 1998: series MB6 (no 2).** *Office for National Statistics* 2000.
24. Hippisley-Cox J, Hammersley V, Pringle M, Coupland C, Crown N, Wright L: **How useful are General Practice databases for research? Analysis of their accuracy and completeness in one research network.** *Health Informatics Journal* 2004, **10**:91-109.
25. Westlake S, Office for National Statistics: **Report: Cancer incidence and mortality in the United Kingdom and constituent countries, 2003-2005.** *HSQ 40* 2008.
26. Valuck R, Williams S, MacArthur M, Saseen J, Nair K, McCollum M, Ensor J: **A retrospective cohort study of correlates of response to pharmacologic therapy for hyperlipidemia in members of a managed care organization.** *Clinical Therapeutics* 2003, **25**:2936-2957.
27. Stattin P, Björ O, Ferrari P, Lukanova A, Lenner P, Lindahl B, Hallmans G, Kaaks R: **Prospective Study of Hyperglycemia and Cancer Risk.** *Diabetes Care* 2007, **30**:561-567.
28. Thomas E, Brewster DH, Black RJ, Macfarlane GJ: **Risk of malignancy among patients with rheumatic conditions.** *International Journal of Cancer* 2000, **88**:497-502.
29. Bangalore S, Kumar S, Kjeldsen SE, Makani H, Grossman E, Wetterslev J, Gupta AK, Sever PS, Gluud C, Messerli FH: **Antihypertensive drugs and risk of cancer: network meta-analyses and trial sequential analyses of 324,168 participants from randomised trials.** *The Lancet Oncology* 2011, **12**:65-82.
30. Henderson KD, Bernstein L: **Etiology of Cancer: Obesity and Physical Activity.** In *DeVita, Hellman, and Rosenberg's cancer: Principles & Practice of Oncology. Volume 1.* Wolters Kluwer/Lippincott Williams & Wilkins; 2008:239-244.
31. Rowan S, Office for National Statistics: **Trends in cancer incidence by deprivation, England and Wales, 1990-2002.** *National Statistics* Health Statistics Quarterly; 2007, 24-35.

32. Gonzalez-Perez A, Garcia Rodriguez L, Lopez-Ridaura R: **Effects of non-steroidal anti-inflammatory drugs on cancer sites other than the colon and rectum: a meta-analysis.** *BMC Cancer* 2003, **3**:28.
33. Sørensen HT, Friis S, Nørgård B, Mellemkjær L, Blot WJ, McLaughlin JK, Ekbom A, Baron JA: **Risk of cancer in a large cohort of nonaspirin NSAID users: a population-based study.** *British Journal of Cancer* 2003, **88**:1687-1692.
34. Heiss G, Wallace R, Anderson GL, Aragaki A, Beresford SAA, Brzyski R, Chlebowski RT, Gass M, LaCroix A, Manson JE, *et al*: **Health Risks and Benefits 3 Years After Stopping Randomized Treatment With Estrogen and Progestin.** *JAMA* 2008, **299**:1036-1045.
35. Royston P: **Multiple imputation of missing values.** *Stata Journal* 2004, **4**:227-241.
36. Royston P: **Multiple imputation of missing values: update of ice.** *The Stata Journal* 2005, **5**:527-536.
37. **Cancer Statistics Registration 2003: series MB1 (no 34).** *Office for National Statistics* 2005.
38. Coogan P, Rosenberg L, Strom B: **Statin use and the risk of 10 cancers.** *Epidemiology* 2007, **18**:213-219.
39. Friedman GD, Flick ED, Udaltsova N, Chan J, Quesenberry CP, Habel LA: **Screening statins for possible carcinogenic risk: up to 9 years of follow-up of 361 859 recipients.** *Pharmacoepidemiology and Drug Safety* 2008, **17**:27-36.
40. Yang YX, Hennessy S, Propert K, Hwang WT, Sarkar M, Lewis JD: **Chronic statin therapy and the risk of colorectal cancer.** *Pharmacoepidemiology and Drug Safety* 2008, **17**:869-876.
41. Singh H, Mahmud SM, Turner D, Xue L, Demers AA, Bernstein CN: **Long-Term Use of Statins and Risk of Colorectal Cancer: A Population-Based Study.** *Am J Gastroenterol* 2009, **104**:3015-3023.
42. Coogan PF, Smith J, Rosenberg L: **Statin Use and Risk of Colorectal Cancer.** *J Natl Cancer Inst* 2007, **99**:32-40.
43. Hoffmeister M, Chang-Claude J, Brenner H: **Individual and joint use of statins and low-dose aspirin and risk of colorectal cancer: A population-based case-control study.** *International Journal of Cancer* 2007, **121**:1325-1330.
44. Flick ED, Habel LA, Chan KA, Haque R, Quinn VP, Van Den Eeden SK, Sternfeld B, Orav EJ, Seeger JD, Quesenberry CPJ, Caan BJ: **Statin Use and Risk of Colorectal Cancer in a Cohort of Middle-Aged Men in the US: A Prospective Cohort Study.** *Drugs* 2009, **69**:1445-1457.
45. Bonovas S, Filioussi K, Sitaras NM: **Statin use and the risk of prostate cancer: A metaanalysis of 6 randomized clinical trials and 13 observational studies.** *International Journal of Cancer* 2008, **123**:899-904.
46. Bonovas S, Filioussi K, Tsavaris N, Sitaras NM: **Use of Statins and Breast Cancer: A Meta-Analysis of Seven Randomized Clinical Trials and Nine Observational Studies.** *Journal of Clinical Oncology* 2005, **23**:8606-8612.
47. Kuliszkiewicz-Janus M, Małecki R, Mohamed A: **Lipid changes occuring in the course of hematological cancers.** *Cellular & Molecular Biology Letters* 2008, **13**:465-474.
48. Sassano A, Katsoulidis E, Antico G, Altman JK, Redig AJ, Minucci S, Tallman MS, Platanias LC: **Suppressive Effects of Statins on Acute Promyelocytic Leukemia Cells.** *Cancer Res* 2007, **67**:4524-4532.
49. Chapman-Shimshoni D, Yuklea M, Radnay J, Shapiro H, Lishner M: **Simvastatin induces apoptosis of B-CLL cells by activation of mitochondrial caspase 9.** *Experimental Hematology* 2003, **31**:779-783.
50. Hippisley-Cox J, Pringle M, Cater R, Wynn A, Hammersley V, Coupland C, Hapgood R, Horsfield P, Teasdale S: **The electronic record in primary care - regression or progression? Cross-sectional survey.** *BMJ* 2003, **326**:1439-1443.
51. Department of Health: **NHS repeat dispensing schemes in England.** 2007.
52. Filion KB, Delaney JAC, Brophy JM, Ernst P, Suissa S: **The impact of over-the-counter simvastatin on the number of statin prescriptions in the United Kingdom: a view from the General Practice Research Database.** *Pharmacoepidemiology and Drug Safety* 2007, **16**:1-4.

Pre-publication history
The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2407/11/409/prepub

doi:10.1186/1471-2407-11-409
Cite this article as: Vinogradova *et al.*: Exposure to statins and risk of common cancers: a series of nested case-control studies. *BMC Cancer* 2011 **11**:409.

Paper 3

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to cyclo-oxygenase-2 inhibitors and risk of cancer: nested case-control studies. *British Journal of Cancer*. 2011;105:452-9.

British Journal of Cancer (2011) 105, 452–459

www.bjcancer.com

Exposure to cyclooxygenase-2 inhibitors and risk of cancer: nested case–control studies

Y Vinogradova*[,1], C Coupland[1] and J Hippisley-Cox[1]

[1]*Division of Primary Care, 13th Floor, Tower Building, University Park, University of Nottingham, Nottingham, NG7 2RD, UK*

BACKGROUND: Selective cyclooxygenase-2 (COX2) inhibitors are widely used as analgesics and it is unclear whether its long-term use affects cancer risk.

METHODS: A series of nested case–control studies using the QResearch primary care database. Associations of COX2 inhibitor use with risk of all cancers and 10 common site-specific cancers were estimated using conditional logistic regression adjusted for comorbidities, smoking status, socioeconomic status, and use of non-steroidal anti-inflammatory drugs, aspirin and statins.

RESULTS: A total of 88 125 cancers, diagnosed between 1998 and 2008, matched with up to five controls, were analysed. Use of COX2 inhibitors for more than a year was associated with a significantly increased risk of breast cancer (odds ratio (OR) 1.24, 95% confidence interval (CI) 1.08–1.42) and haematological malignancies (OR 1.38, 95% CI 1.12–1.69) and a decreased risk of colorectal cancer (OR 0.76, 95% CI 0.63–0.92). There were no other significant associations.

CONCLUSION: Prolonged use of COX2 inhibitors was associated with an increased risk of breast and haematological cancers and decreased risk of colorectal cancer. These findings need to be confirmed using other data sources.

British Journal of Cancer (2011) **105,** 452–459. doi:10.1038/bjc.2011.252 www.bjcancer.com
Published online 12 July 2011

Keywords: COX2 inhibitors; breast cancer; colorectal cancer; haematological cancers; QResearch

Selective cyclooxygenase-2 (COX2) inhibitors are used for patients intolerant to traditional non-steroidal anti-inflammatory drugs (NSAID), which have gastrointestinal toxic effects (Chan *et al*, 2010). Being introduced in the United Kingdom in 1985, COX2 inhibitors account for 14% of all NSAIDs prescriptions (The NHS Information Centre for health and social care, 2008), despite advice from the UK Medicines and Healthcare products Regulatory Agency (MHRA, 2005) about possible cardiovascular adverse effects (Solomon *et al*, 2005).

Laboratory investigations have suggested mechanisms by which COX2 inhibitors might reduce the risk of cancer (Koki and Masferrer, 2002; Khan and Lee, 2009) for a range of cancers, although animal experiments have not provided consistent support. A recent publication, for example, shows that COX2 inhibitors do not delay or prevent tumour development in breast tissue in a mouse model (Tran-Thanh *et al*, 2010).

Some observational studies have investigated effects of COX2 inhibitors on cancer risk, but have produced inconsistent results (Arber *et al*, 2006; Harris *et al*, 2006, 2007; Hernández-Díaz and García Rodríguez, 2006). For colorectal cancer, a randomised control trial (Arber *et al*, 2006) showed a 36% decreased rate of newly detected colorectal adenomas in celecoxib users. Two studies (Harris *et al*, 2006, 2007) demonstrated risk reductions for breast and lung cancer, but a larger case–control study (Hernández-Díaz and García Rodríguez, 2006) using primary care data showed no effect for lung cancer. Effects on other cancers remain unclear.

We designed a series of large-scale nested case–control studies to determine associations between selective COX2 inhibitors and risks of common cancers. We used the QResearch primary care database, which is large, has a representative population and contains data for individual drug exposures and outcomes.

MATERIALS AND METHODS

Study design, data source and population

We conducted a series of nested case–control studies using version 20 of the QResearch primary care database (http://www.qresearch.org) containing anonimised clinical records for over 11 million patients registered with 574 UK general practices. The information recorded on the database includes patient demographics (year of birth, sex, sociodemographic data derived from the UK census 2001), characteristics (height, weight, smoking status), clinical diagnoses, symptoms, consultations, referrals, prescribed medications and results of investigations. The database has been validated by comparing birth rates, death rates, consultation rates, prevalence and mortality rates with other data sources, including the General Household Survey and the General Practice Research Database (National Statistics, 2000; Hippisley-Cox *et al*, 2005).

We initially identified an open cohort of patients registered between 1 Jan 1997 and 1 July 2008 with participating UK general practices. We then selected as cases all those patients in the cohort aged between 30 and 100 years with a first-ever recorded diagnosis of cancer during the study period, identified from diagnostic

*Correspondence: Dr Y Vinogradova;
E-mail: yana.vinogradova@nottingham.ac.uk
Received 10 March 2011; revised 2 June 2011; accepted 10 June 2011; published online 12 July 2011

READ codes in patient records (the standard clinical terminology system used in General Practice in the UK (Smith *et al*, 1995)). Each case was linked to five controls who were alive, had no history of cancer and were registered with the practice at the time of case diagnosis (the index date), matched on age, sex, practice and calendar time using incidence-density sampling.

Exclusions

Cases with secondary cancers (READ codes: B56, B57, B58) and non-melanoma skin cancer were excluded. For breast cancer, we included only females, and excluded cases and controls with a record of mastectomy or tamoxifen use for more than 12 months before the index date to exclude possible previous diagnoses. We also excluded temporary residents and patients with fewer than 6 years of medical records before the index date to ensure completeness of exposure data.

Primary outcomes

We analysed cancers overall, and carried out separate analyses for the most common UK cancers (Westlake, 2008): breast (women, B34), prostate (men, B46), lung (B22), colorectal (B13, B14), haematological (B6), bladder (B49), melanoma (B32), gastric (B11), pancreatic (B17) and oesophageal (B10). As haematological malignancies cover a range of diseases, possibly differentially affected by COX2 inhibitors (Nakanishi *et al*, 2001; Nakamura *et al*, 2006), we also investigated leukaemia (B63-B6z), lymphoma (B60-B62) and myeloma (B63) separately.

Data

Records in the year before the index date were ignored to reduce protopathic bias. Prescriptions for cases in this period could relate to early cancer symptoms before the recorded diagnosis. All analyses were, therefore, based only on prescriptions relating to the period between 13 and 72 months before the index date.

We assessed exposure to COX2 inhibitors, including celecoxib, etodolac, etoricoxib, lumiracoxib, rofecoxib, valdecoxib, meloxicam (British Medical Association, Royal Pharmaceutical Society, 2010). We also extracted data on prescriptions for statins, traditional NSAIDs and aspirin because studies have found protective effects of these on various types of cancer (Garcia Rodriguez and Huerto-Alvarez, 2001; Sørensen *et al*, 2003; Jacobs *et al*, 2005; Bardia *et al*, 2007; Gallicchio *et al*, 2007), in particular, colorectal cancer (Garcia Rodriguez and Huerto-Alvarez, 2001; Sørensen *et al*, 2003).

We extracted information on age and sex, smoking status (non-smoker, ex-smoker, current smoker), body mass index (BMI) in $kg\,m^{-2}$, Townsend score (measure of socioeconomic status) and data on comorbidities (cardiovascular disease, hypertension, diabetes, rheumatoid arthritis and osteoarthritis). For breast cancer, we also accounted for previous benign breast disease (fibrocystic disease, intraductal papilloma, fibroadenoma), family history of breast cancer, use of hormone replacement therapy and oral contraceptives. For colorectal cancer, additional comorbidities were ulcerative colitis and Crohn's disease.

We considered patients as COX2 inhibitor users if they had at least one prescription. We estimated cumulative use of COX2 inhibitors by extracting the duration of use for every prescription and, for groups of prescriptions with inter-prescription gaps of less than 60 days; we calculated overall course times from the start of the first prescription to the end of the last prescription. We then calculated cumulative use as the sum of all overall course times and categorised cumulative use for each patient as: no use, less than 90 days, 90 days to 12 months; 13–24 months; 25–60 months. We also categorised cumulative use as: no use; short-term use (less than 365 days) and long-term use (more than 365 days). A trend test was performed using the actual months of use. We conducted separate analyses for the most common individual COX2 inhibitors – meloxicam, rofecoxib and celecoxib, examining the effect on cancer risk of cumulative use for more than 365 days.

The daily dose of COX2 inhibitors was estimated as the median daily dose of all prescriptions of any COX2 drug recorded. It was categorised by COX2 inhibitor efficacy (Hernández-Díaz and García Rodríguez, 2006) as: high (for celecoxib >200 mg, for meloxicam >7.5 mg, for rofecoxib >25 mg, for etodolac >400 mg, for etoricoxib >90 mg, for valdecoxib >40 mg, for luminoracoxib >200 mg); otherwise as low/medium.

The effect on cancer risk of stopping COX2 inhibitors for long-term and short-term users was investigated by determining the last prescription date and categorising each patient at 12 months before the index date as: no COX2 inhibitors use, current COX2 inhibitors user, recent user (stopped the drugs at 13–24 months before the index date) and past user (stopped the drugs at 25 or more months before the index date).

Statistical analysis

We used conditional multivariate logistic regression to estimate odds ratios (ORs) with 95% confidence intervals (CIs) associated with COX2 inhibitor use compared with non-use for cancers overall and each specific cancer. We calculated unadjusted ORs and adjusted for the potential confounding variables listed above, in which patients were classified as users of each medication if they had at least one prescription for NSAIDs or aspirin and at least two prescriptions for statins, hormone replacement therapy and oral contraceptives.

We carried out multiple imputation (Royston, 2005) with Stata ICE programs to replace missing values of BMI, smoking status and Townsend deprivation scores. We applied Rubin's rules to five imputed data sets to combine effect estimates for each cancer separately. We removed rheumatoid arthritis patients in an additional analysis to eliminate its potential effect on the risk of haematological malignancies Thomas *et al*, 2000).

We used all the available data on the QResearch database, hence, did not do a pre-study sample size calculation. We chose a 1% significance level to determine statistical significance to account for the multiple outcomes. Stata v10 (StataCorp LP, College Station, TX, USA) was used for all analyses.

RESULTS

There were 118 780 patients with diagnoses of cancers in the study period matched with 588 797 controls. Of the patients with cancer, 3810 with secondary cancers and 36 with inapplicable cancers (e.g., male/cervical cancer) were removed. For breast cancer, 1055 cases and 773 controls with a previous mastectomy or tamoxifen use were excluded. This left 113 879 cases with a first diagnosis of cancer during the study period and 568 958 matched controls. After removing 25 754 cases and 206 704 controls with <6 years of medical records or lacking a matched case or control, there were 88 125 cases of primary cancer matched with 362 254 controls, which were used in the analyses. The proportions of each cancer type in cases matched registration statistics in England for 2007 (Statistical Bulletin, 2010) for patients older than 30 years.

Baseline characteristics

Table 1 shows baseline characteristics for cases and controls. Fifty-three percent of cases were men; with a median age at diagnosis of 69 years (interquartile range: 60–77). Overall, 76% of cases and 73% of controls had complete data for BMI, smoking status and Townsend deprivation score. Cases and controls had similar patterns of comorbidity.

Table 1 Baseline characteristics for all cases with primary cancer and their matched controls with at least 6 years of medical records

	Cases ($N = 88\,125$)	Controls ($N = 362\,254$)
Sex		
Female	41 749 (47.4)	170 173 (47.0)
Male	46 376 (52.6)	192 081 (53.0)
Age band (years)		
30–54	13 151 (14.9)	49 906 (13.8)
55–64	19 638 (22.3)	80 107 (22.1)
65–74	26 758 (30.4)	111 698 (30.8)
75–84	25 013 (28.4)	106 278 (29.3)
85+	3565 (4.0)	14 265 (3.9)
Deprivation, Townsend quintile		
1, Most affluent	22 072 (25.0)	92 287 (25.5)
2	18 998 (21.6)	79 067 (21.8)
3	17 338 (19.7)	71 358 (19.7)
4	15 325 (17.4)	61 767 (17.1)
5, Most deprived	11 896 (13.5)	45 971 (12.7)
Townsend missing	2496 (2.8)	11 804 (3.3)
Body mass index ($kg\,m^{-2}$)		
15–24	26721 (30.3)	105 883 (29.2)
25–29	27 285 (31.0)	108 803 (30.0)
30–49	12 922 (14.7)	51 413 (14.2)
Not recorded	21 197 (24.1)	96 155 (26.5)
Smoking status		
Non-smoker	54 307 (61.6)	233 135 (64.4)
Ex-smoker	7567 (8.6)	23 842 (6.6)
Current smoker	17 275 (19.6)	54 869 (15.1)
Not recorded	8976 (10.2)	50 408 (13.9)
Comorbidities		
Cardiovascular disease	14 278 (16.2)	58 123 (16.0)
Diabetes	7115 (8.1)	26 802 (7.4)
Hypertension	27 104 (30.8)	109 797 (30.3)
Osteoarthritis	12 807 (14.5)	52 586 (14.5)
Rheumatoid arthritis	1310 (1.5)	5132 (1.4)
Colitis[a]	124 (1.1)	293 (0.6)
Crohn's disease[a]	28 (0.2)	109 (0.2)
Benign breast disease[b]	1094 (7.0)	2937 (4.7)
Family history of breast cancer[b]	539 (3.4)	1249 (2.0)
Medications (in previous 13–72 months)		
Traditional NSAIDs	35 697 (40.5)	140 642 (38.8)
Aspirin	19 895 (22.6)	79 067 (21.8)
Statins	13 621 (15.5)	54 606 (15.1)
Hormone replacement therapy[b]	3289 (21.0)	10 973 (17.4)
Oral contraceptive pill[b]	523 (3.3)	1638 (2.6)

Abbreviation: NSAID = non-steroidal anti-inflammatory drug. [a]On the basis of cases with colorectal cancer and their controls only. [b]On the basis of female cases with breast cancer and their controls only. Values are shown as numbers and %.

Exposure to COX2 inhibitors

Overall 7.8% (6901) of cases and 7.4% (26 974) of controls had at least one prescription for COX2 inhibitors. Most users (70% cases, 70% controls) had no gap longer than 60 days between the first and last prescription, with 19% cases and 19% controls having only one gap longer than 60 days. Twenty-one percent of COX2 inhibitor users (21% cases, 21% controls) had prescriptions for more than 365 days (Figure 1) with median 20 prescriptions (interquartile range, 14–30 for cases, 14–31 for controls). Median duration of use for these long-term users was 25 months (interquartile range 17–37 for cases and 18–37 for controls) and median duration for short-term users was 2 months (interquartile range 1–4 for both cases and controls).

The most frequently prescribed COX2 inhibitors were rofecoxib (3.1% cases, 3.0% controls), celecoxib (2.6% cases, 2.5% controls) and meloxicam (2.6% cases, 2.4% controls). Other COX2 inhibitors were prescribed to <1% cases and 1% controls. Most rofecoxib users were on low/medium dose (71% cases, 72% controls), most celecoxib users were on high dose (81% cases, 79% controls) and more than half of meloxicam users (65% cases, 65% controls) were on high dose.

A higher proportion of COX2 inhibitors users had hypertension, cardiovascular disease, diabetes, rheumatoid arthritis and osteoarthritis than non-users (Table 2).

Cancer of any site The analysis for cancer risk showed a significant association with any COX2 inhibitors use, although the OR (Table 3) was close to unity (OR 1.06, 95% CI 1.03–1.09, $P < 0.001$), and no association for long-term use (OR 1.02, 95% CI 0.96–1.08, $P = 0.616$). Analyses of trends for duration of use and dosage, as well as individual COX2 inhibitors use did not show significant associations with overall cancer risk (Table 4 and Supplementary information).

Colorectal cancer There was no association between any use of COX2 inhibitors and risk of colorectal cancer, but the association with long-term use was significant (OR 0.76, 95% CI 0.63–0.92, $P = 0.004$). There was a significant trend for duration of use ($P_{trend} = 0.004$) with an OR of 0.66 (95% CI 0.51–0.86, $P = 0.002$) for more than 24 months of use. Risk of colorectal cancer stayed significantly decreased for long-term users who stopped COX2 inhibitors more than 2 years before the index date (OR 0.74, 95% CI 0.60–0.92, $P = 0.007$).

Breast cancer Risk of breast cancer was not statistically significantly associated with overall COX2 inhibitor use, but there was a significant trend with duration of use ($P_{trend} = 0.002$) with an increased risk in long-term users (OR 1.24, 95% CI 1.08–1.42, $P = 0.003$), which stayed increased after stopping COX2 inhibitors more than 2 years before the index date (OR 1.23, 95% CI 1.05–1.44, $P = 0.009$).

Haematological malignancies There was a significant association between risk of haematological malignancies and COX2 inhibitor use (OR 1.18, 95% CI 1.07–1.31, $P = 0.001$) with an even stronger association for long-term users (OR 1.38, 95% CI 1.12–1.69, $P = 0.002$). There was a significant trend for duration of use ($P_{trend} < 0.001$) and an increased risk of 47% in users for more than 2 years ($P = 0.008$). Removing cases and controls with rheumatoid arthritis did not change the ORs. Meloxicam had the highest OR (OR 1.27, 95% CI 1.08–1.50, $P = 0.004$) for overall use, but others were not statistically significant. The risk in long-term users remained significantly increased after stopping COX2 inhibitors for more than 2 years before the index date (OR 1.40, 95% CI 1.11–1.76, $P = 0.005$).

The ORs for overall use in separate analyses for leukaemia, lymphoma and myeloma showed consistent increases, though only myeloma was significant (ORs 1.18, 95% CI 1.02–1.36, $P = 0.030$; 1.21, 95% CI 1.01–1.45, $P = 0.036$; and 1.43, 95%CI 1.13–1.81. $P = 0.003$, respectively). Long-term use showed a stronger effect for lymphoma (ORs 1.20, 95% CI 0.88–1.64, $P = 0.246$; 1.70, 95% CI 1.21–2.40, $P = 0.002$; and 1.38, 95% CI 0.87–2.19, $P = 0.168$, for leukaemia, lymphoma and myeloma, respectively), with respective trends ($P_{trend} = 0.071$), ($P_{trend} = 0.001$) and ($P_{trend} = 0.048$) for actual months of use.

Lung cancer There were no significant associations for lung cancer. Long-term COX2 inhibitor users had a lower risk (OR 0.79, 95% CI 0.65–0.95, $P = 0.012$), but it was not statistically significant at the level of 0.01.

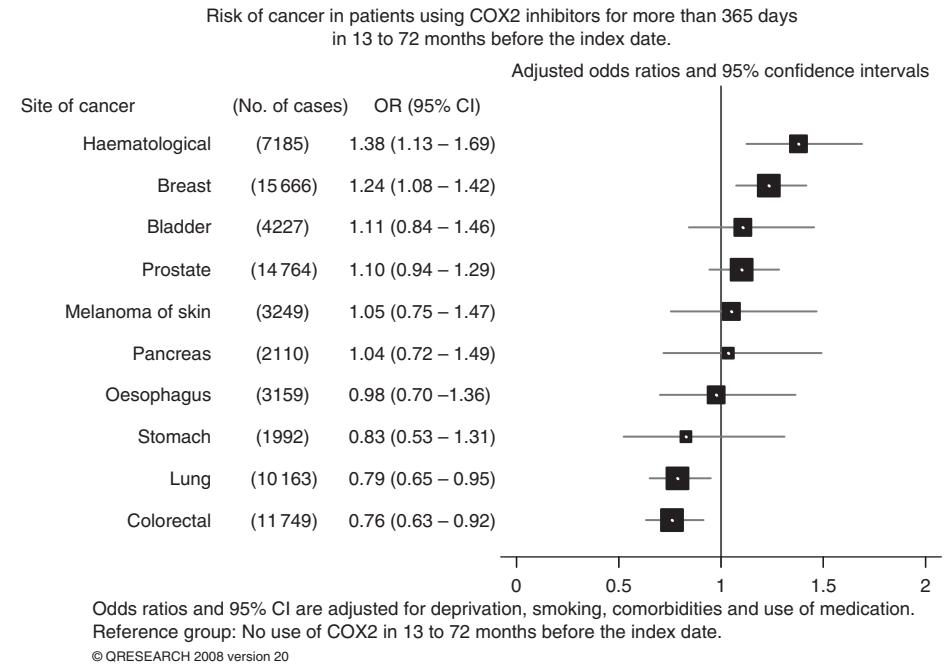

Figure 1 Risk of cancer in patients using COX2 inhibitors for more than 365 days in 13–72 months before the index date.

Other cancers There were no significant associations with COX2 inhibitor use for other cancers.

Other analyses No dose–response association with cancer was found for any site. No particular type of COX2 inhibitor overall use was associated with increased or decreased risk of cancer (except for blood cancer reported above).

DISCUSSION

The key findings from our study are that long-term use of selective COX2 inhibitors was associated with a 24% reduced risk of colorectal cancer, a 24% increased risk of breast cancer and a 38% increased risk of haematological cancer. No significant increases or decreases for other common cancers were found. Although the protective effect for colorectal cancer might have been hypothesised from theoretical and laboratory studies (Koki and Masferrer, 2002; Khan and Lee, 2009), we believe this is the first demonstration using general population clinical data.

Comparison with other studies

Many epidemiological studies have investigated the effects of nonspecified or combined (COX2 and traditional) NSAIDs on cancer risk (Garcia Rodriguez and Huerto-Alvarez, 2001; Sørensen *et al*, 2003; Jacobs *et al*, 2005; Hernández-Díaz and García Rodríguez, 2006; Bardia *et al*, 2007; Gallicchio *et al*, 2007). A number of them have suggested overall chemoprotective properties of NSAIDs for several cancers, in particular colorectal (Garcia Rodriguez and Huerto-Alvarez, 2001; Sørensen *et al*, 2003) and, for long-duration regular users, lung, prostate and breast cancer (Jacobs *et al*, 2005; Hernández-Díaz and García Rodríguez, 2006; Gallicchio *et al*, 2007).

There is less evidence for newer COX2 drugs, although laboratory and animal studies (Liu *et al*, 2004; Manish *et al*, 2005; Barnes *et al*, 2007; D'Arca *et al*, 2010) using COX2 inhibitors have shown possible decreases in cancer incidence. The reduced risk of colorectal cancer in our study was comparable with the 56% decreased risk of distal large bowel cancer in COX2 inhibitor users (Kim *et al*, 2008). COX2 inhibitor chemoprotective effects were also demonstrated in a randomised controlled trial for colorectal cancer prevention (Arber *et al*, 2006), although on patients with increased baseline risk because of previous history of adenomas. Although the trial was planned for 5 years of surveillance and treatment, it was stopped after 3.1 years because of adverse cardiovascular effects, but it still demonstrated a significant anti-tumour effect with risk reductions of 55–67% depending on celecoxib dose (Bertagnolli *et al*, 2006).

Our study's finding of an increased risk of breast cancer contrasts with findings from a hospital-based case–control study on selective COX-2 inhibitors (Harris *et al*, 2006), which demonstrated a significant risk reduction (OR 0.29, 95% CI 0.14–0.59) with daily use for at least 2 years. This study was very small (only 10 cases), and used questionnaire data and hence would have been subject to recall bias. Another study (Rahme *et al*, 2005) on menopausal women showed a reduction in breast cancer risk (OR 0.81, 95% CI 0.68–0.97) for COX2 inhibitor use of 90 days or longer, however, with shorter exposure (average of eight prescriptions). Although no other recent epidemiological study has looked at specific effects of COX-2 inhibitors, a number of studies have investigated effects of nonspecified or combined NSAID use on breast cancer (Gill *et al*, 2007; Kirsh *et al*, 2007; Ready *et al*, 2008), mostly finding no association. The mechanism of inhibiting of COX2 expression might differ for different types of traditional NSAIDs, and a cohort study (Marshall *et al*, 2005) demonstrated an increased risk in ibuprofen users but not in aspirin or other NSAIDs.

We showed increased risks for haematological malignancies, particularly lymphoma. Frequent traditional NSAID users with rheumatoid arthritis may have double the risk of having haematological cancers (Thomas *et al*, 2000), and one rheumatoid arthritis study showed an increased risk of lymphoma (Baecklund *et al*, 2006) from chronic inflammation. Removing rheumatoid arthritis patients, left our results unchanged, suggesting an effect from COX2 inhibitors rather than from the condition. Another meta-analysis demonstrated no association between NSAIDs and non-Hodgkin lymphoma (Bernatsky *et al*, 2007) risk, but the only study on COX2 inhibitors found a possible increased risk

Table 2 Baseline characteristics in cases and controls COX2 users and non-users (at least one prescription in 13 to 72 months before index date)

	Cases ($N = 88\,125$)		Controls ($N = 362\,254$)	
	COX2 user $N = 6901$	**COX2 non-users** $N = 81\,224$	**COX2 user** $N = 26\,974$	**COX2 non-users** $N = 335\,280$
Sex				
Female	3833 (55.5)	37 916 (46.7)	15 055 (55.8)	155 118 (46.3)
Male	3068 (44.5)	43 308 (53.3)	11 919 (44.2)	180 162 (53.7)
Age band (years)				
30–54	455 (6.6)	12 696 (15.6)	1693 (6.3)	48 213 (14.4)
55–64	1309 (19.0)	18 329 (22.6)	4830 (17.9)	75 277 (22.5)
65–74	2123 (30.8)	24 635 (30.3)	8616 (31.9)	103 082 (30.7)
75–84	2466 (35.7)	22 547 (27.8)	9911 (36.7)	96 367 (28.7)
85+	548 (7.9)	3017 (3.7)	1924 (7.1)	12 341 (3.7)
Deprivation, Townsend quintile				
1, Most affluent	1606 (23.3)	20 466 (25.2)	6378 (23.6)	85 909 (25.6)
2	1467 (21.3)	17 531 (21.6)	5876 (21.8)	73 191 (21.8)
3	1344 (19.5)	15 994 (19.7)	5491 (20.4)	65 867 (19.6)
4	1334 (19.3)	13 991 (17.2)	4933 (18.3)	56 834 (17.0)
5, Most deprived	955 (13.8)	10 941 (13.5)	36 39 (13.5)	42 332 (12.6)
Townsend missing	195 (2.8)	2301 (2.8)	657 (2.4)	11 147 (3.3)
Body mass index ($kg\,m^{-2}$)				
15–24	1802 (26.1)	24 919 (30.7)	7228 (26.8)	98 655 (29.4)
25–29	2491 (36.1)	24 794 (30.5)	9686 (35.9)	99 117 (29.6)
30–49	1482 (21.5)	11 440 (14.1)	5814 (21.6)	45 599 (13.6)
Not recorded	1126 (16.3)	20 071 (24.7)	4246 (15.7)	91 909 (27.4)
Smoking status				
Non-smoker	4659 (67.5)	49 648 (61.1)	19 789 (73.4)	213 346 (63.6)
Ex-smoker	742 (10.8)	6825 (8.4)	2346 (8.7)	21 496 (6.4)
Current smoker	1215 (17.6)	16 060 (19.8)	3661 (13.6)	51 208 (15.3)
Not recorded	285 (4.1)	8691 (10.7)	1178 (4.4)	49 230 (14.7)
Comorbidities				
Cardiovascular disease	1483 (21.5)	12 795 (15.8)	5991 (22.2)	52 132 (15.5)
Diabetes	711 (10.3)	6404 (7.9)	2601 (9.6)	24 201 (7.2)
Hypertension	2858 (41.4)	24 246 (29.9)	11 449 (42.4)	98 348 (29.3)
Osteoarthritis	2667 (38.6)	10 140 (12.5)	10 623 (39.4)	41 963 (12.5)
Rheumatoid arthritis	400 (5.8)	910 (1.1)	1519 (5.6)	3613 (1.1)
Colitis[a]	8 (0.9)	116 (1.1)	34 (0.9)	259 (0.6)
Crohn's disease[a]	2 (0.2)	26 (0.2)	9 (0.2)	100 (0.2)
Benign breast disease[b]	92 (7.1)	1002 (7.0)	210 (4.2)	2727 (4.7)
Family history of breast cancer[b]	37 (2.8)	502 (3.5)	88 (1.7)	1161 (2.0)
Medications (in previous 13–72 months)				
Traditional NSAIDs	1771 (25.7)	11 850 (14.6)	6971 (25.8)	47 635 (14.2)
Aspirin	4629 (67.1)	31 068 (38.2)	18 229 (67.6)	122 413 (36.5)
Statins	2290 (33.2)	17 605 (21.7)	9047 (33.5)	70 020 (20.9)
Hormone replacement therapy[b]	324 (24.8)	2965 (20.6)	1102 (21.8)	9871 (17.1)
Oral contraceptive pill[b]	13 (1.0)	510 (3.6)	50 (1.0)	1588 (2.7)
Medications in the last 12 months				
COX2 inhibitors	1754 (25.4)	1763 (2.2)	7136 (26.5)	5341 (1.6)

Abbreviations: COX2 = cyclooxygenase-2; NSAID = non-steroidal anti-inflammatory drug. [a]On the basis of cases with colorectal cancer and their controls, only. [b]On the basis of female cases with breast cancer and their controls, only. Values are shown as numbers and %.

associated with regular use (Flick *et al*, 2006) (OR 1.58, 95% CI 0.68–3.67). A recent study (Chang *et al*, 2010) also demonstrated an increased risk of Hodgkin lymphoma, associated with COX2 inhibitors.

There is no established biological mechanism explaining the associations between COX2 inhibitors and risk of breast or blood cancers, and further exploration is needed.

We found no significant reduction of lung cancer risk in patients with over 1 year use of COX2 inhibitors, although there was some indication of a decreased risk (OR 0.79, 95% CI 0.65–0.95), in contrast to a very small study reporting a 60% reduction for COX2 inhibitor use of 2 years or more (Harris *et al*, 2007) (22 cases) with inevitable recall bias. A larger case–control study demonstrated a reduction of risk (Hernández-Díaz and García Rodríguez, 2006), based on all NSAIDS, but no significant association for COX2 inhibitors.

Strengths and limitations

The study was substantially larger than earlier studies, including information from all patients, including those with short survival. There is no recall bias, as details of prescriptions and confounding

Table 3 Use of selective COX2 inhibitors (at least one prescription) in cases and in controls in 13 to 72 months before the index date by cancer site

Cancer	Total number of cases	Total number of controls	*No.* of COX2 inhibitors users in cases (%)	*No.* of COX2 inhibitors users in controls (%)	Unadjusted OR (95% CI)[a]	Adjusted OR (95% CI)[a,b]	Adjusted *P*-value
Breast[c]	15 666	62 938	1304 (8.3)	5046 (8.0)	1.09 (1.02–1.17)	1.07 (1.00–1.15)	0.047
Prostate	14 764	61 853	1067 (7.2)	3979 (6.4)	1.16 (1.08–1.24)	1.09 (1.01–1.18)	0.022
Colorectal[d]	11 749	48 624	866 (7.4)	3752 (7.7)	0.97 (0.90–1.05)	0.99 (0.91–1.08)	0.817
Lung	10 163	42 415	845 (8.3)	3500 (8.3)	1.03 (0.95–1.12)	1.00 (0.91–1.09)	0.922
Haematological	7185	29 162	634 (8.8)	2104 (7.2)	1.30 (1.18–1.44)	1.18 (1.07–1.31)	0.001
Bladder	4227	17 559	332 (7.9)	1239 (7.1)	1.17 (1.03–1.34)	1.15 (1.00–1.32)	0.045
Skin	3249	13 115	239 (7.4)	952 (7.3)	1.06 (0.91–1.24)	1.05 (0.89–1.23)	0.579
Oesophagus	3159	13 041	222 (7.0)	941 (7.2)	0.99 (0.85–1.16)	1.03 (0.88–1.21)	0.710
Pancreas	2110	8762	189 (9.0)	716 (8.2)	1.11 (0.94–1.33)	1.12 (0.94–1.35)	0.215
Stomach	1992	8279	143 (7.2)	573 (6.9)	1.07 (0.87–1.30)	1.03 (0.84–1.27)	0.747
All cancers	88 125	362 254	6901 (7.8)	26 974 (7.4)	1.09 (1.06–1.12)	1.06 (1.03–1.09)	<0.001

Abbreviations: CI = confidence interval; COX2 = cyclooxygenase-2; OR = odds ratio. [a]Compared with no use. [b]Adjusted for Townsend quintile, body mass index, smoking status, myocardial infarction, coronary heart disease, diabetes, hypertension, stroke, rheumatoid arthritis, osteoarthritis, use of other lipid-lowering drugs, non-steroidal anti-inflammatory drugs, COX2 inhibitors and aspirin. [c]Also adjusted for family history of breast cancer, use of oral contraceptives and hormone replacement therapy. [d]Also adjusted for colitis and Crohn's disease.

Table 4 Cumulative duration of COX2 inhibitors use in cases and controls in 13–72 months before the index date by cancer site

	Less than 90 days		90 days–12 months		13–24 months		25 months and more		
Cancer	Cases/ controls	Adjusted odds ratio (95% CI)[a]	Cases/ controls	Adjusted odds ratio (95% CI)[a]	Cases/ controls	Adjusted odds ratio (95% CI)[a]	Cases/ controls	Adjusted odds ratio (95% CI)[a]	*P*-value[b]
Breast[c]	684/2888	0.98 (0.89–1.07)	329/1143	1.21 (1.06–1.37)[d]	144/482	1.29 (1.07–1.57)[d]	147/533	1.19 (0.98–1.44)	0.002
Prostate	621/2287	1.11 (1.01–1.22)	225/869	1.04 (0.90–1.21)	109/398	1.11 (0.90–1.38)	112/425	1.09 (0.88–1.36)	0.097
Colorectal[e]	514/2062	1.07 (0.96–1.18)	207/873	1.04 (0.88–1.21)	80/401	0.88 (0.69–1.12)	65/416	0.66 (0.51–0.86)[d]	0.004
Lung	506/1905	1.10 (0.98–1.23)	175/762	0.95 (0.79–1.14)	62/385	0.60 (0.45–0.81)[d]	102/448	0.96 (0.75–1.22)	0.138
Haematological	320/1170	1.09 (0.95–1.25)	169/527	1.25 (1.04–1.50)	72/212	1.31 (0.99–1.74)	73/195	1.47 (1.11–1.95)[d]	<0.001
Bladder	190/662	1.25 (1.05–1.49)	69/297	0.99 (0.75–1.30)	32/139	1.01 (0.68–1.51)	41/141	1.20 (0.84–1.72)	0.369
Skin	137/529	1.06 (0.87–1.30)	54/223	1.00 (0.73–1.36)	32/90	1.49 (0.98–2.27)	16/110	0.68 (0.39–1.16)	0.524
Oesophagus	132/537	1.09 (0.88–1.33)	43/199	0.96 (0.68–1.35)	23/115	0.87 (0.55–1.39)	24/90	1.11 (0.70–1.76)	0.465
Pancreas	97/383	1.07 (0.84–1.36)	53/168	1.37 (0.99–1.90)	18/83	0.95 (0.56–1.61)	21/82	1.08 (0.66–1.78)	0.583
Stomach	86/314	1.13 (0.87–1.46)	33/138	1.01 (0.68–1.51)	14/58	1.01 (0.55–1.84)	10/63	0.67 (0.34–1.32)	0.262
All cancers	3884/15 021	1.07 (1.03–1.11)	1602/6161	1.08 (1.02–1.15)[d]	693/2806	1.03 (0.95–1.12)[d]	722/2986	1.00 (0.92–1.09)	0.236

Abbreviations: CI = confidence interval; COX2 = cyclooxygenase-2. [a]Adjusted for Townsend quintile, body mass index, smoking status, myocardial infarction, coronary heart disease, diabetes, hypertension, stroke, rheumatoid arthritis, osteoarthritis, use of other lipid-lowering drugs, non-steroidal anti-inflammatory drugs, COX2 inhibitors, aspirin and compared with no use. [b]Trend test based on number of months prescribed. [c]Also adjusted for family history of breast cancer, use of oral contraceptives, hormone replacement therapy. [d]*P*-value < 0.01. [e]Also adjusted for colitis and Crohn's disease.

factors were recorded prospectively before the index date. Bias from misclassification of diagnoses was unlikely because accuracy and completeness of records in general practices is high (Hippisley-Cox *et al*, 2003; Herrett *et al*, 2010). Matching controls on sex, age, practice and calendar year removed effects from these confounding factors and we adjusted for a number of other confounding variables. Although we used a 1% level to define statistical significance level, some of our findings might still have arisen from multiple significance testing. Bias from misclassification of COX2 inhibitor use was unlikely as over 99% of all repeat prescriptions are computer recorded (Department of Health, 2007), and underestimation of use was unlikely as these drugs are prescription-only.

We did not adjust for certain cancer risk factors, such as physical activity, women's reproductive history, alcohol use and diet, because these are not consistently recorded. There may, therefore, be residual confounding if these factors are associated with COX2 inhibitor use. Body mass index, smoking status and deprivation had missing values in 22% of cases and in 25% of controls, and we used multiple imputation to replace these values. Although our data contain detailed information on drug prescriptions, this may not reflect the actual use. There is no reason to think that any non-adherence would systematically differ between cases and controls, however, such misclassification might have biased the ORs towards one making the associations weaker. There may be residual confounding because of over-the-counter use of NSAIDs and aspirin, which was not accounted for in the analyses. There was no information about cancer stage and it is unknown whether the symptoms before diagnosis led to COX2 inhibitor use. The possibility of this was minimised by ignoring prescriptions in the last year before the index date.

Summary

We have conducted a large population-based case–control study examining the association of selective COX-2 inhibitors with risk of common cancers in the general population and found a reduced risk of colorectal cancer, but increased risks of breast and haematological malignancies in long-term COX2 inhibitor users, which did not decrease after cessation. This was a very broad study covering a range of cancers, each of which, though related, are complex and exhibit significant variations in terms of disease mechanisms and progression, symptoms and treatments. The primary value of the study is, therefore, as a comprehensive overview, identifying the relative potential of different areas for further focused investigation. Although some significant findings are reported, further studies are suggested, in particular, in the areas of breast and blood cancers.

ACKNOWLEDGEMENTS

We acknowledge the contribution of EMIS for expertise in creating and maintaining QResearch and to the EMIS practices, which contribute data, without whom this research would not be possible.

Author contributions

YV contributed to the study design, undertook the literature review and the primary analysis as well as the first interpretation and wrote the first draft of the paper. YV is the guarantor of the study. CC contributed to the development of the idea, design, analysis, interpretation and drafting of the paper. JH-C had the original idea for this study, extracted the data, contributed to design, interpretation and commented on the draft of the paper.

Conflict of interest

JH-C is co-director of QResearch (a not for profit organization, that is, a joint partnership between the University of Nottingham and EMIS, the leading commercial supplier of IT for 60% of general practices in the United Kingdom) and director of ClinRisk, which produces software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to improve patient care. This work and any views expressed within it are solely those of the authors and not of any affiliated bodies or organisations.

Supplementary Information accompanies the paper on British Journal of Cancer website (http://www.nature.com/bjc)

REFERENCES

Arber N, Eagle CJ, Spicak J, Rácz I, Dite P, Hajer J (2006) Celecoxib for the prevention of colorectal adenomatous polyps. *N Engl J Med* **355:** 885 – 895

Baecklund E, Iliadou A, Askling J, Ekbom A, Backlin C, Granath F, Catrina AI, Rosenquist R, Feltelius N, Sundström C, Klareskog L (2006) Association of chronic inflammation, not its treatment, with increased lymphoma risk in rheumatoid arthritis. *Arthritis Rheum* **54**(3)**:** 692 – 701

Bardia A, Ebbert J, Vierkant R, Limburg P, Anderson K, Wang A, Olson J, Vachon C, Cerhan J (2007) Association of aspirin and nonaspirin nonsteroidal anti-inflammatory drugs with cancer incidence and mortality. *J Natl Cancer Inst* **99**(11)**:** 881 – 889

Barnes NLP, Warnberg F, Farnie G, White D, Jiang W, Anderson E, Bundred NJ (2007) Cyclooxygenase-2 inhibition: effects on tumour growth, cell cycling and lymphangiogenesis in a xenograft model of breast cancer. *Br J Cancer* **96**(4)**:** 575 – 582

Bernatsky S, Lee JL, Rahme E (2007) Non-Hodgkin's lymphoma–meta-analyses of the effects of corticosteroids and non-steroidal anti-inflammatories. *Rheumatology* **46**(4)**:** 690 – 694

Bertagnolli MM, Eagle CJ, Zauber AG, Redston M, Solomon SD, Kim K, Tang J, Rosenstein RB, Wittes J, Corle D, Hess TM, Woloj GM, Boisserie F, Anderson WF, Viner JL, Bagheri D, Burn J, Chung DC, Dewar T, Foley TR, Hoffman N, Macrae F, Pruitt RE, Saltzman JR, Salzberg B, Sylwestrowicz T, Gordon GB, Hawk ET (2006) Celecoxib for the prevention of sporadic colorectal adenomas. *N Engl J Med* **355**(9)**:** 873 – 884

British Medical Association, Royal Pharmaceutical Society (2010) *British National Formulary*, vol. 59. British Medical Journal Publishing Group and Pharmaceutical Press: London

Chan FKL, Lanas A, Scheiman J, Berger MF, Nguyen H, Goldstein JL (2010) Celecoxib versus omeprazole and diclofenac in patients with osteoarthritis and rheumatoid arthritis (CONDOR): a randomised trial. *Lancet* **376**(9736)**:** 173 – 179

Chang E, Cronin-Fenton D, Friis S, Hjalgrim H, Sørensen H, Pedersen L (2010) Aspirin and other nonsteroidal anti-inflammatory drugs in relation to Hodgkin lymphoma risk in northern Denmark. *Cancer Epidemiol Biomarkers Prev* **19**(1)**:** 59 – 64

D'Arca D, LeNoir J, Wildemore B, Gottardo F, Bragantini E, Shupp-Byrne D, Zanesi N, Fassan M, Croce CM, Gomella LG, Baffa R (2010) Prevention of urinary bladder cancer in the FHIT knock-out mouse with Rofecoxib, a Cox-2 inhibitor. *Urol Oncol* **28**(2)**:** 189 – 194

Department of Health (2007) *NHS Repeat Dispensing Schemes In England*. Department of Health: London

Flick ED, Chan KA, Bracci PM, Holly EA (2006) Use of nonsteroidal antiinflammatory drugs and non-Hodgkin lymphoma: a population-based case-control study. *Am J Epidemiol* **164**(5)**:** 497 – 504

Gallicchio L, Visvanathan K, Burke A, Hoffman SC, Helzlsouer KJ (2007) Nonsteroidal anti-inflammatory drugs and the risk of developing breast cancer in a population-based prospective cohort study in Washington County, MD. *Int J Cancer* **121:** 211 – 215

Garcia Rodriguez LA, Huerto-Alvarez C (2001) Reduced risk of colorectal cancer among long-term users of asprin and nonaspirin nonsteroidal antiinflammatory drugs. *Epidemiology* **12**(1)**:** 88 – 93

Gill JK, Maskarinec G, Wilkens LR, Pike MC, Henderson BE, Kolonel LN (2007) Nonsteroidal antiinflammatory drugs and breast cancer risk: the multiethnic cohort. *Am J Epidemiol* **166**(10)**:** 1150 – 1158

Harris RE, Beebe-Donk J, Alshafie GA (2006) Reduction in the risk of human breast cancer by selective cyclooxygenase-2 (cox-2) inhibitors. *BMC Cancer* **6:** 27

Harris RE, Beebe-Donk J, Alshafie GA (2007) Reduced risk of human lung cancer by selective cyclooxygenase 2 (cox-2) blockade: results of a case control study. *Int J Biol Sci* **3:** 328 – 334

Hernández-Díaz S, García Rodríguez L (2006) Nonsteroidal anti-inflammatory drugs and risk of lung cancer. *Int J Cancer* **120:** 1565 – 1572

Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ (2010) Validation and validity of diagnoses in the general practice research database: a systematic review. *Br J Clin Pharmacol* **69**(1)**:** 4 – 14

Hippisley-Cox J, Pringle M, Cater R, Wynn A, Hammersley V, Coupland C, Hapgood R, Horsfield P, Teasdale S (2003) The electronic record in primary care - regression or progression? Cross-sectional survey. *BMJ* **326:** 1439 – 1443

Hippisley-Cox J, Vinogradova Y, Coupland C, Pringle M (2005) *Comparison of key practice characteristics between general practices in England and Wales and general practices in the QResearch data. Report to the Health and Social Care Information Centre*. University of Nottingham. Report no. R 21

Jacobs E, Rodriguez C, Mondul A, Connell C, Henley S, Calle E, Thun M (2005) A large cohort study of aspirin and other nonsteroidal anti-inflammatory drugs and prostate cancer incidence. *J Natl Cancer Inst* **97:** 975 – 980

Khan MNA, Lee YS (2009) Cyclooxygenase inhibitors: scope of their use and development in cancer chemotherapy. *Med Res Rev* **31**(2)**:** 161 – 201

Kim S, Martin C, Galanko J, Woosley JT, Schroeder JC, Keku TO, Satia JA, Halabi S, Sandler RS (2008) Use of nonsteroidal antiinflammatory drugs and distal large bowel cancer in whites and African Americans. *Am J Epidemiol* **168**(11)**:** 1292 – 1300

Kirsh VA, Kreiger N, Cotterchio M, Sloan M, Theis B (2007) Nonsteroidal antiinflammatory drug use and breast cancer risk: subgroup findings. *Am J Epidemiol* **166**(6)**:** 709 – 716

Koki A, Masferrer J (2002) Celecoxib: a specific COX-2 inhibitor with anticancer properties. *Cancer Control* **9**(2 Suppl)**:** 28 – 35

Liu X, Yue P, Zhou Z, Khuri F, Sun S (2004) Death receptor regulation and celecoxib-induced apoptosis in human lung cancer cells. *J Natl Cancer Inst* **96**(23)**:** 1769 – 1780

Manish I, Patel MI, Subbaramaiah K, Du B, Chang M, Yang P, Newman RA, Cordon-Cardo C, Thaler HT, Dannenberg AJ (2005) Celecoxib inhibits prostate cancer growth: evidence of a cyclooxygenase-2-independent mechanism. *Clin Cancer Res* **11:** 1999 – 2007

Marshall SF, Bernstein L, Anton-Culver H, Deapen D, Horn-Ross PL, Mohrenweiser H, Peel D, Pinder R, Purdie DM, Reynolds P, Stram D, West D, Wright WE, Ziogas A, Ross RK (2005) Nonsteroidal anti-inflammatory drug use and breast cancer risk by stage and hormone receptor status. *J Natl Cancer Inst* **97**(11)**:** 805 – 812

MHRA (2005) *Updated Advice On the Safety of Selective Cox-2 Inhibitors*. Medicines and Healthcare products Regulatory Agency: London

Nakamura S, Kobayashi M, Shibata K, Sahara N, Shigeno K, Shinjo K, Naito K, Hayashi H, Ohnishi K (2006) Etodolac induces apoptosis and inhibits cell adhesion to bone marrow stromal cells in human myeloma cells. *Leuk Res* **30**(2)**:** 123 – 135

Nakanishi Y, Kamijo R, Takizawa K, Hatori M, Nagumo M (2001) Inhibitors of cyclooxygenase-2 (COX-2) suppressed the proliferation and differentiation of human leukaemia cell lines. *Eur J Cancer* **37**(12)**:** 1570 – 1578

National Statistics (2000) *Key Health Statistics from General Practice 1998: Series MB6 (no 2)*. Office for National Statistics: London

Rahme E, Ghosn J, Dasgupta K, Rajan R, Hudson M (2005) Association between frequent use of nonsteroidal anti-inflammatory drugs and breast cancer. *BMC Cancer* **5**(1)**:** 159

Ready A, Velicer C, McTiernan A, White E (2008) NSAID use and breast cancer risk in the VITAL cohort. *Breast Cancer Res Treat* **109**(3)**:** 533 – 543

Royston P (2005) Multiple imputation of missing values: update of ice. *Stata J* **5**(4)**:** 527 – 536

Smith N, Wilson A, Weekes T (1995) Use of Read codes in development of a standard data set. *BMJ* **311:** 313 – 315

Solomon SD, McMurray JJV, Pfeffer MA, Wittes J, Fowler R, Finn P, Anderson WF, Zauber A, Hawk E, Bertagnolli M (2005) Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. *N Engl J Med* **352**(11)**:** 1071 – 1080

Sørensen HT, Friis S, Nørgård B, Mellemkjær L, Blot WJ, McLaughlin JK, Ekbom A, Baron JA (2003) Risk of cancer in a large cohort of nonaspirin NSAID users: a population-based study. *Br J Cancer* **88:** 1687 – 1692

Statistical Bulletin (2010) *Cancer Statistics Registration: Registrations of Cancer Diagnosed in 2007, England: Series MB1 (no 38)*. Office for National Statistics: London

The NHS Information Centre for health and social care (2008) *Prescription Cost Analysis, England 200*. Prescribing Support Unit: London

Thomas E, Brewster DH, Black RJ, Macfarlane GJ (2000) Risk of malignancy among patients with rheumatic conditions. *Int J Cancer* **88**(3)**:** 497 – 502

Tran-Thanh D, Buttars S, Wen Y, Wilson C, Done S (2010) Cyclooxygenase-2 inhibition for the prophylaxis and treatment of preinvasive breast cancer in a her-2/neu mouse model. *Cancer Prev Res* **3**(2)**:** 202 – 211

Westlake S, Office for National Statistics (2008) Report: cancer incidence and mortality in the United Kingdom and constituent countries, 2003 – 2005. HSQ 40

Paper 4

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of cancer: a protocol for nested case-control studies using the QResearch primary care database. *BMJ Open*. 2012;2(1):e000548.

BMJ open accessible medical research

Exposure to bisphosphonates and risk of cancer: a protocol for nested case–control studies using the QResearch primary care database

Yana Vinogradova, Carol Coupland, Julia Hippisley-Cox

To cite: Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of cancer: a protocol for nested case–control studies using the QResearch primary care database. *BMJ Open* 2012;**2**: e000548. doi:10.1136/bmjopen-2011-000548

► Prepublication history for this paper is available online. To view these files please visit the journal online (http://bmjopen.bmj.com).

Received 27 October 2011
Accepted 25 November 2011

Division of Primary Care, University of Nottingham, School of Community Health Sciences, Division of Primary Care, University Park, Nottingham, UK

Correspondence to
Yana Vinogradova; yana.vinogradova@nottingham.ac.uk

ABSTRACT

Introduction: Bisphosphonates are becoming a common treatment for osteoporosis particularly after discovery of the association between hormone replacement therapy and increased risk of breast cancer. As osteoporosis develops with age, treatment is a long-term intervention. Randomised control trials typically have limited follow-up times, which restricts investigation of the effects of the drugs on risk of primary cancers. A few observational studies have demonstrated a reduced risk of breast cancer and possibly of endometrial cancer in bisphosphonate users. Two epidemiological studies have studied the effect of the drugs on oesophageal cancer but did not reach any definite conclusions. So far, no effects on colorectal and stomach cancer have been shown. This study will investigate the association of bisphosphonates with risks of the 10 most common primary cancers.

Methods and analysis: A series of nested case–control studies will be based on the general population using records from 660 UK general practices within the QResearch Database. Cases will be patients with primary cancers diagnosed between 1996 and 2011. Each case will be matched by age, sex, practice and calendar year to five controls, who are alive and registered with the practice at the time of diagnosis of the case. Exposure to bisphosphonates will be defined as at least one prescription during the study period. For the most common cancers with substantial numbers of observations, the effect of the duration of the treatment and different types of bisphosphonates will be studied. Conditional logistic regression will be applied to produce ORs adjusted for smoking status, socioeconomic status, ethnicity, cancer-specific co-morbidities and use of other medications.

ARTICLE SUMMARY

Article focus

- Bisphosphonate use.
- Effect on incidence of cancer.
- Designing a study.

Key messages

- Series of case–control studies will examine possible associations between use of bisphosphonates and risk of cancer.
- Effect of dose, duration and different types of drug will be investigated.
- Results will be adjusted for a number of confounders.

Strengths and limitations of this study

- Large sample size.
- Based on the general populations.
- Based on routinely collected data.
- Prescriptions not actual use.

INTRODUCTION

Osteoporosis among the older people is a major problem leading to increased mortality and morbidity and high costs for health services. Thirty-five per cent of the European population aged 50 years and over suffer from fractures caused by osteoporosis.[1] Between 1980 and 1990, the use of hormone replacement therapy (HRT) was considered a preventive measure for postmenopausal osteoporotic fractures in women but, after a Women's Health Initiative trial report about increased risk of breast cancer, use of HRT fell significantly.[2]

As a treatment for postmenopausal osteoporosis, bisphosphonates were introduced in the 1990s, and prescribing of them has increased substantially and continually. HRT (raloxifene) and the use of calcitonin and strontium ranelate[3] are still considered to be options for the treatment of osteoporosis, but according to the UK National Institute for Health and Clinical Excellence guidelines,[4 5] recommending bisphosphonates as a first-line therapy for osteoporosis bisphosphonates have become the most commonly prescribed drug.

The proportion of the female population in the UK eligible for treatment varies

between 24% and 47%, depending on age.[6] The drugs increase bone mass and reduce the risk of fracture, but these effects become significant only after 6–36 months of use depending on the type of drug.[7] Bisphosphonates bind to bone and, depending on type, can be released for up to ten more years after treatment ceases.[8]

The first use of bisphosphonates in the 1970s was in oncology. They were used for the treatment and prevention of skeletal disorders associated with multiple myeloma and bone metastases from breast, prostate, lung and kidney cancers and other solid tumours. Bisphosphonates have also been used for glucocorticoid-induced osteoporosis.[7]

There is preclinical evidence for the anti-tumour effects of bisphosphonates because of their anti-resorptive properties.[9] Bone is a good environment for tumour cells because of a number of growth factors. Osteoclasts affect release of soluble growth factors and so promote tumour cells. Bisphosphonates accumulated in bones inhibit osteoclast-mediating bone resorption with significant clinical effect. The drugs also demonstrate anti-tumour effects in vitro by inhibiting angiogenesis (adhesion, invasion and proliferation) and inducing apoptosis. The cancers studied in vitro were breast, prostate, myeloma, pancreatic and osteosarcoma.[10] These preclinical studies, however, were conducted with concentrations far higher than those used for treating patients with bone metastates.[11]

Although the anti-tumour properties of bisphosphonates are being considered for prevention of bone metastases and a few clinical trials have demonstrated the efficacy of bisphosphonates in women with early-stage breast cancer,[12] they have been little studied in relation to the development of other primary cancers. Four epidemiological studies concentrating on breast cancer have shown positive effects for bisphosphonates: 32% RR reduction in postmenopausal women (HR 0.68, 95% CI 0.52 to 0.88),[13] 33% decreased risk in current users, women aged 20–69 years (OR 0.67, 95% CI 0.51 to 0.89),[14] 39% risk reduction in patients taking bisphosphonates for at least 1 year (OR 0.61, 95% CI 0.50 to 0.76)[15] and 47% risk reduction after start of alendronate (HR 0.53, 95% CI 0.38 to 0.73) and 20% for etidronate (HR 0.80, 95% CI 0.73 to 0.89).[16] A study looking at the risk of endometrial cancer has also shown a 30% decrease associated with bisphosphonate use, but it was not statistically significant (OR 0.7, 95% CI 0.4 to 1.2).[17]

Because bisphosphonates are associated with short-term gastrointestinal adverse effects,[8] an adverse effect on risk of oesophageal cancer might be expected. The first publication about the association was from the US Food and Drug Administration Adverse Event Reporting System, which listed 23 cases of oesophageal cancer in users of oral alendronate between 1995 and 2008.[18] A further observational study, based on 13 678 bisphosphonate users matched to 27 365 non-users, identified 37 oesophageal cancers and 48 gastric cancers and showed reduced risks for oesophageal and gastric cancers (HR 0.35, 95% CI 0.14 to 0.85 and HR 1.23, 95% CI 0.68 to 2.22, respectively).[19]

A case–control study looking at 2954 cases of oesophageal, 2018 cases of gastric and 10 641 cases of colorectal cancers, based on the General Practice Research Database, demonstrated a 30% increased risk of oesophageal cancer in patients with at least one prescription for bisphosphonates (OR 1.30, 95% CI 1.02 to 1.66)[20] but did not find a significant effect on risk of gastric or colorectal cancers (OR 0.87, 95% CI 0.64 to 1.19 and OR 0.87, 95% CI 0.77 to 1.00, respectively). A cohort study based on the General Practice Research Database did not find any significant association between bisphosphonate use and risk of gastric or oesophageal cancers[21] (combined HR 0.96, 95% CI 0.74 to 1.25, for oesophageal cancer only HR 1.07, 95% CI 0.77 to 1.49). As for colorectal cancer, an Israeli study showed a significantly decreased risk in patients taking bisphosphonates for more than a year (RR 0.50, 95% CI 0.25 to 0.67).[22] A Danish study looked at gastrointestinal cancers and reported an excess risk of oesophageal cancer associated with use of alendronate (HR 2.10, 95% CI 1.01 to 4.35) and etidronate (HR 1.99, 95% CI 1.24 to 3.18) and a possible protective effect of higher doses for colorectal cancer (HR 0.29, 95% CI 0.14 to 0.62).[23] So far no epidemiological studies have investigated associations with risks of other common cancers for bisphosphonate users. A few randomised controlled trials—the longest for up to 10 years[24–26]—have studied the effect of the drugs on skeletal properties and general adverse effects, but none of them have considered cancer as a consequence of osteoporotic therapy. A cohort study in patients treated for osteoporosis including bisphosphonates is currently enrolling participants to explore a number of adverse events in the next 5 years.[27] This is the only study where malignancies form part of the secondary outcome measures.

Our aim is to examine possible associations between use of bisphosphonates and risk of a range of common cancers in a large community sample, including the effect of dose, duration and type of drug.

METHODS AND ANALYSIS

Sample selection

This will be a study using the QResearch primary care research database, which consists of routinely collected data from general practitioner clinical computer systems. The contributing practices, which comprise around 7% of all UK general practices, use the Egton Medical Information System. QResearch is one of the largest general practice databases, containing anonymised clinical records for over 13 million patients registered with 660 UK general practices. The information recorded on the database includes patient demographics (year of birth, sex, socio-demographic data derived from UK census 2001), characteristics (height, weight, smoking status), clinical diagnoses, symptoms

and prescribed medications (including repeat prescriptions). Detailed analyses, including age and sex distribution, birth rates and death rates, have been undertaken and have shown good correspondence with other sources[28] and demonstrated the accuracy and completeness of the data.[29]

An open cohort of patients will be identified, 30 years or older, registered with the study practices during the study period, between 1 January 1996 and 1 July 2011. Temporary residents will be excluded. Cases will be incident cases of cancer identified during the study period, and these will include the 10 most common cancers. Cases with any previous cancer diagnosis will be excluded. Cases with secondary cancers (READ codes: B56, B57, B58) will be excluded. The right censor date will be the earliest of the following: date of diagnosis of cancer, date of death, date of leaving the practice, date of the latest download of data, the study end date.

Cases and controls

Each case will be matched to five controls, who are alive and registered with the practice at the time of diagnosis of the case. Controls will be matched on age, sex, practice and calendar year using incidence density sampling. Controls will be allocated an index date, which is the date on which their matched case was first diagnosed with cancer. Controls with a diagnosis of any cancer before the index date will be excluded.

Cases and controls with a record of mastectomy before their first prescription of bisphosphonates will be excluded since this treatment is likely to indicate a previous diagnosis of breast cancer with further bone metastases. For breast cancer, only female patients will be included. All patients with Paget's disease will be excluded as the treatment for this condition is administered in higher doses and for much longer periods (typically 2 weeks for osteoporosis against 6 months for Paget's). Patients with prescriptions for the bisphosphonates licensed not for osteoporosis but for malignancies (zoledronic acid, clodronate and daily use of ibandronate) will also be excluded.

For the main analysis, cases and controls will be included if they have complete records for at least 2 years before the index date. A subset of cases and controls with at least 6 years of records will be used for further analyses.

The risks of any cancer and of the 10 most common cancers will be determined for patients prescribed bisphosphonates and compared with the risks for patients not prescribed these drugs. The 'most common' cancers have been selected because they have this status in the UK.[30] They are breast cancer (women, B34), prostate cancer (men, B46), lung cancer (B22), colorectal cancer (B13, B14), haematological malignancies (B6), bladder cancer (B49), melanoma (B32), gastric cancer (B11), pancreatic cancer (B17) and oesophageal cancer (B10). As osteoporosis might be an early symptom of possible myeloma, it will be analysed separately from lymphoma and leukaemia. The commoner female cancers (ovary (B44), uterus (B43) and cervix (B41)) will also be considered.

Interventions

Exposure to drugs for osteoporosis will be determined based on all prescriptions for bisphosphonates and other drugs before the index date (date of diagnosis or equivalent date for controls) within the observation period (from the date of entry into QResearch to the index date). The bisphosphonates to be included are identified in the British National Formulary section 6.6.2 as treatment for osteoporosis[3]: alendronate (5−10 mg daily or 70 mg weekly), etidronate (400 mg daily for 14 days in 90-day cycles), ibandronate (150 mg a month or intravenous 3 mg/3 months) and risedronate (5mg daily).

The cumulative exposure to bisphosphonates will be assessed by extracting duration of the prescribed days' supply and summarising it for each patient. For drugs prescribed in cycles, the length of a cycle will be considered as duration of a prescription, for example, etidronate prescription for 2 weeks will be assessed as a 90-day prescription duration. The same approach will be applied to intravenous infusion, considering the recommended interval between injections as the duration of a prescription (eg, 3 months for ibandronate). The cumulative exposure to bisphosphonates will be estimated by extracting the duration for every prescription, and for groups of prescriptions with inter-prescription gaps of <60 days, overall course times will be calculated from the start of the first prescription to the end of the last prescription.

As bisphosphonates can be released for months after a treatment, total exposure to bisphosphonates will be estimated as the time between the first prescription and the end time for the last prescription.

Because bisphosphonates and other osteoporosis treatment drugs are prescribed for years, long-term users and short-term users will be distinguished, as treatment of the latter might have been for accidental or clinical fractures or for better integration of biomaterial or implants. The effect of bisphosphonates on treating fractures varies from 6 to 36 months, for example 12 months for risedronate and 24 months for alendronate.[7]

There are three regimens for bisphosphonate use: daily, once weekly and once monthly. Daily use has been shown to have lower adherence than weekly use.[31] Another reason for investigating regimens is that, particularly for gastrointestinal organs, there might be a marked difference between the effects of daily and weekly exposure to bisphosphonates, with associated effects on risks for oesophageal, gastric and colorectal cancers.

Bisphosphonate use will be categorised in a number of ways. The main analyses will compare patients having no prescriptions for the drugs with patients with at least one prescription for any bisphosphonate. The effect of prescribing for short-term (<12 months) and long-term (at least 12 months) periods will then be analysed, as well as the effect of regimen: daily or weekly/monthly.

If there are a sufficient number of observations, further analyses will be run for the cumulative exposure (cumulative duration of all prescriptions) and the exposure time to bisphosphonate (the time period between the first prescription and the end time for the last prescription). The subset of data with at least 6 years of records will be analysed using following categorisations: no use, <180 days, 180 days up to 12 months, 12–24 months and ≥25 months. A test for trend will be performed using the actual number of months.

Timing will be categorised as: no use before diagnosis, used within 1–2 years before the index date and used >2 years before the index date. The interaction of timing and terms of treatment will also be examined, categorised as: no use before diagnosis; used within 1–2 years before the index date, short-term use; used within 1–2 years before the index date, long-term use; used >2 years before the index date, short-term use and used >2 years before the index date, long-term use.

If there are any variations in dose of bisphosphonates, it will be categorised as low (<67% of dose recommended by dose) and normal/high (>66% of recommended dose).

The two main types of bisphosphonates—simple bisphosphonates (etidronate) and nitrogen containing[7]—will be analysed as there are two different mechanisms of action for the drugs. If there are sufficient numbers, the data will also be analysed by individual drug.

Because prescriptions in the year before the index date might be associated with an early symptom of cancer before a recorded diagnosis, sensitivity analyses ignoring all prescriptions in the last year before the index date will be run. The results from these analyses will highlight any attenuation of the protective effects of bisphosphonates or any increases in magnitudes of harmful effects. A sensitivity analysis on the main analysis will also be run, defining the use of bisphosphonates as at least two prescriptions within the observation period. The analyses will be repeated on a subgroup of patients with at least 6 years of records to estimate the long-term effect of bisphosphonate use.

The other drugs for osteoporosis to be included are strontium ranelate, raloxifene and calcitonin. As there will not be enough observations to analyse each drug individually, they will be combined and included in the analyses as other treatment for osteoporosis. A patient will be considered as a user if they have at least one prescription of any of those drugs in their records before the index date.

Confounding factors

All the analyses will include potential confounders which are established as risk factors for cancer: body mass index[32] (continuous variable, at the date closest to 1 year before the diagnosis and recorded before the index date); smoking status[33] (current smoker—light (1–9 cigarettes/day), medium (10–19 cigarettes/day) and heavy (≥20 cigarettes/day); ex-smoker and non-smoker); excessive alcohol consumption[34] using Read codes for alcohol status (only if it is a significant confounder for the sample); socioeconomic status[35] (Townsend score in fifths) and ethnicity[36] (Caucasian, African–American, Asian and other). The analysis will also adjust for osteoporosis history,[37] including diagnosis of osteoporosis or osteopenia or previous fractures, use of drugs increasing risk of fracture (systemic corticosteroids and proton pump inhibitors[38]), use of anti-inflammatory drugs[39] (traditional non-steroidal, cyclooxygenase 2 inhibitors and aspirin)[40] and use of vitamin D.[41]

Co-morbidities which affect risks of cancer will also be included: rheumatoid arthritis[42] for any cancer; hypertension[43] for uterine cancer and diabetes and glucose intolerance for pancreatic,[44] uterine[45] and colorectal[46] cancers. Analyses of colorectal, oesophageal, gastric and pancreatic cancers will be adjusted for gastrointestinal disorders[47] if diagnosed before the first use of bisphosphonates or 12 months before the index date, whichever is earlier: upper gastrointestinal disease (dysphagia, oesophagitis, gastrooesophageal reflux disease, hiatus hernia, oesophageal ulcers, Barrett's oesophagus, gastritis, duodenitis, peptic ulcers, dyspepsia); Crohn's disease, ulcerative colitis, and pancreatitis. Bladder cancer analyses will include renal impairment[48] (diagnostic code for chronic kidney disease) if diagnosed before the first use of bisphosphonates or 12 months before the index date, whichever is earlier. Breast cancer analyses will also include previous benign breast disease (fibrocystic disease, intraductal papilloma or fibroadenoma).[49] The results will also be adjusted for family history of cancer[50] (this will vary according to the cancer under consideration) if recorded 6 months before the index date. This is to reduce family recall bias as cases are more likely to report a family history of cancer around the time of diagnosis.[51]

Because use of some drugs might be associated with increased risk of some cancers, use of HRT[52] and oral contraceptives[53] for breast, uterine, ovarian and cervical cancers will also be included. Use of acid suppression drugs[54] (including H2 antagonists (BNF 1.3.1), proton pump inhibitors (BNF 1.3.5) and antacids (BNF 1.1.1)) will be added for gastrointestinal cancer analyses. If there are enough observations, use of those drugs will be categorised by the number of prescriptions within the observation period: none, fewer than 12 prescriptions, 12–24 prescriptions, 25–48 prescriptions and >49 prescriptions.

Statistical analysis

Conditional logistic regression will be used to estimate OR with 95% CIs for cancer of any site and each of the 10 most common cancers and three additional female cancers and their matched controls. The initial analysis model will determine the unadjusted ORs for each cancer associated with bisphosphonate prescriptions. A multivariable model will determine the OR for each

cancer associated with bisphosphonate prescriptions, adjusted for the potential confounding effects of the variables listed above.

As body mass index, smoking status and alcohol consumption may be important confounders but have non-negligible numbers of missing data, multiple imputation will be used to impute the missing values. Ten imputed data sets will be created. Index year, case/control status, years of records, potential confounders and exposure to bisphosphonates and other drugs will be included. For comparison, analyses with missing data treated as separate categories will also be carried out.

Stata V.11 will be used for all the analyses. A 1% significance level will be used to account for the multiple outcomes.

Sample size calculation

As different types of cancer may have different risks associated with bisphosphonate use, analyses will require number of cases to relate to each type of cancer. All available data from QResearch will be used. Our calculations are based on the exposure to bisphosphonates in the proposed data extraction for 6.8% of women and 1.8% of men. For non-gender-specific cancers, the total proportion of users is estimated as 4.2%. To detect an OR of 0.87 (for colorectal or stomach cancers[20]), 22 322 cases will be needed. To detect an OR of 1.3 (for oesophageal cancer[20]), 5208 cases will be needed. To detect an OR of 0.70 (for breast[14] or uterus[17] cancers), 2382 female cases will be needed. For prostate cancer, to detect 30% increase (or decrease) in risk, 11 773 (or 8686) male cases will be needed. For other cancers, a detection of 30% risk decrease will require 3785 cases. All calculations are done for matched sets of cases and controls, with 4.5 matched controls per case, an estimated coefficient for exposure between matched cases and controls of 0.2, a power of 80% and a significance level of 1%.

Acknowledgements We acknowledge the contribution of Egton Medical Information System and the University of Nottingham for expertise in creating and maintaining QResearch and to the Egton Medical Information System practices, which contribute data without whom this research would not be possible.

Funding This work has been funded by the Division of Primary Care of University of Nottingham. Apart from that, this research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no support from any additional organisation for the submitted work. JH-C is professor of clinical epidemiology at the University of Nottingham and unpaid director of QResearch, a not-for-profit organisation, which is a joint partnership between the University of Nottingham and Egton Medical Information System (commercial IT supplier for 60% of general practices in the UK). JH-C is also a paid director of ClinRisk Limited, which produces open and closed source software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help improve patient care. CC is associate professor of medical statistics at the University of Nottingham and a paid consultant statistician for ClinRisk Limited; no other relationships or activities that could appear to have influenced the submitted work.

Ethics approval This protocol has been independently peer-reviewed by the QResearch Scientific Board and has been reported to Trent Research Ethics Committee in accordance with the agreed procedure. A full report containing the study findings will be prepared and a paper based on the report will be submitted to a peer-reviewed journal.

Contributors JH-C had the original idea for this study. CC contributed to the development of the idea. YV reviewed the literature, contributed to the study design and wrote the draft of the manuscript. JH-C and CC critically reviewed the paper. YV is the guarantor of the study. All authors approved the submitted version.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement It will be possible to access the data after the publication of the results but only on premises of the University of Nottingham. The full protocol and statistical code will be available from the authors after the publication of the results.

REFERENCES

1. Johnell O, Kanis JA. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporos Int* 2006;17:1726–33.
2. Watson J, Wise L, Green J. Prescribing of hormone therapy for menopause, tibolone, and bisphosphonates in women in the UK between 1991 and 2005. *Eur J Clin Pharmacol* 2007;63:843–9.
3. Joint Formulary Committee. British National Formulary London. *British Medical Association and Royal Pharmaceutical Society of Great Britain*. 62 ed. London: BMJ Group and Pharmaceutical Press; 2011.
4. National Institute for Health and Clinical Excellence. Alendronate, etidronate, risedronate, strontium ranelate and raloxifen for preventing bone fractures in postmenopausal women with osteoporosis who have not had a fracture. information about NICE technology appraisal guidance. London: National Institute for Health and Clinical Excellence, 2008;160.
5. National Institute for Health and Clinical Excellence. Alendronate, etidronate, risedronate, strontium ranelate and teriparatide for preventing bone fractures in postmenopausal women with osteoporosis who have already had a fracture. Information about NICE technology appraisal guidance. 2008;161.
6. Kanis J, McCloskey E, Johansson H, *et al*. Case finding for the management of osteoporosis with FRAX®—assessment and intervention thresholds for the UK. *Osteoporos Int* 2008;19:1395–408.
7. Russell RG, Watts NB, Ebetino FH, *et al*. Mechanisms of action of bisphosphonates: similarities and differences and their potential influence on clinical efficacy. *Osteoporos Int* 2008;19:733–59.
8. Watts NB, Diab DL. Long-Term use of bisphosphonates in osteoporosis. *J Clin Endocrinol Metab* 2010;95:1555–65.
9. Guise TA. Antitumor effects of bisphosphonates: promising preclinical evidence. *Cancer Treat Rev* 2008;34(Suppl 1):S19–24.
10. Croucher P, Jagdev S, Coleman R. The anti-tumor potential of zoledronic acid. *Breast* 2003;12(Suppl 2):S30–6.
11. Santini D, Fratto ME, Galluzzo S, *et al*. Are bisphosphonates the suitable anticancer drugs for the elderly? *Crit Rev Oncol Hematol* 2009;69:83–94.
12. Gnant M. Can oral bisphosphonates really reduce the risk of breast cancer in Healthy women? *J Clin Oncol* 2010;28:3548–51.
13. Chlebowski RT, Chen Z, Cauley JA, *et al*. Oral bisphosphonate use and breast cancer incidence in postmenopausal women. *J Clin Oncol* 2010;28:3582–90.
14. Newcomb PA, Trentham-Dietz A, Hampton JM. Bisphosphonates for osteoporosis treatment are associated with reduced breast cancer risk. *Br J Cancer* 2010;102:799–802.
15. Rennert G, Pinchev M, Rennert HS. Use of bisphosphonates and risk of postmenopausal breast cancer. *J Clin Oncol* 2010;28:3577–81.
16. Vestergaard P, Fischer L, Mele M, *et al*. Use of bisphosphonates and risk of breast cancer. *Calcif Tissue Int* 2011;88:255–62.
17. Fortuny J, Sima C, Bayuga S, *et al*. Risk of endometrial cancer in relation to medical conditions and medication use. *Cancer Epidemiol Biomarkers Prev* 2009;18:1448–56.
18. Wysowski DK. Reports of esophageal cancer with oral bisphosphonate use. *N Engl J Med* 2009;360:89–90.
19. Abrahamsen B, Eiken P, Eastell R. More on reports of esophageal cancer with oral bisphosphonate use [letter]. *N Engl J Med* 2009;360:1789.
20. Green J, Czanner G, Reeves G, *et al*. Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort. *BMJ* 2010;341:c4444.

21. Cardwell CR, Abnet CC, Cantwell MM, *et al*. Exposure to oral bisphosphonates and risk of esophageal cancer. *JAMA* 2010;304:657–63.
22. Rennert G, Pinchev M, Rennert HS, *et al*. Use of bisphosphonates and reduced risk of colorectal cancer. *J Clin Oncol* 2011;29:1146–50.
23. Vestergaard P. Occurrence of gastrointestinal cancer in users of bisphosphonates and other antiresorptive drugs against osteoporosis. *Calcif Tissue Int* 2011;89:434–41.
24. Mellström DD, Sörensen OH, Goemaere S, *et al*. Seven years of treatment with risedronate in women with postmenopausal osteoporosis. *Calcif Tissue Int* 2004;75:462–8.
25. Black DM, Schwartz AV, Ensrud KE, *et al*. Effects of continuing or stopping alendronate after 5 Years of treatment: the Fracture Intervention Trial Long-term Extension (FLEX): a randomized trial. *JAMA* 2006;296:2927–38.
26. Bone HG, Hosking D, Devogelaer JP, *et al*. Ten years' experience with alendronate for osteoporosis in postmenopausal women. *N Engl J Med* 2004;350:1189–99.
27. Pfizer. *Bazedoxifene Post Approval Safety Study (PASS) in the European Union (EU)*. 2011. B1781044. http://www.clinicaltrials.gov
28. Hammersley V, Hippisley-Cox J, Wilson A, *et al*. A comparison of research general practices and their patients with other practices—cross sectional survey in Trent. *Br J Gen Pract* 2002;52:463–8.
29. Hippisley-Cox J, Hammersley V, Pringle M, *et al*. How useful are General Practice databases for research? Analysis of their accuracy and completeness in one research network. *Health Inform J* 2004;10:91–109.
30. Westlake S. Cancer incidence and mortality in the United Kingdom and constituent countries, 2004-06. *Health Stat Q*. 2009 Autumn; 43:56–62.
31. Recker RR, Gallagher R, MacCosbe PE. Effect of dosing frequency on bisphosphonate medication adherence in a large longitudinal cohort of women. *Mayo Clin Proc* 2005;80:856–61.
32. Henderson KD, Bernstein L. *Etiology of Cancer: Obesity and Physical Activity. DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology*. Wolters Kluwer/Lippincott Williams & Wilkins, Philadelphia, 2008:239–44.
33. Hecht SS. *Etiology of Cancer: Tobacco. DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology*. Wolters Kluwer/ Lippincott Williams & Wilkins, Philadelphia, 2008:147–55.
34. Schütze M, Boeing H, Pischon T, *et al*. Alcohol attributable burden of incidence of cancer in eight European countries based on results from prospective cohort study. *BMJ* 2011;342:d1584.
35. Forman D. *Cancer Incidence by Deprivation. England, 1995-2004*. 2008. http://wwwncinorguk/publications/reports/defaultaspx, http://www.ncin.org.uk/publications/reports/default.aspx
36. Ferlay J, Shin HR, Bray F, *et al*. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127:2893–917.
37. McGlynn KA, Gridley G, Mellemkjaer L, *et al*. Risks of cancer among a cohort of 23,935 men and women with osteoporosis. *Int J Cancer* 2008;122:1879–84.
38. Yang YX, Lewis JD, Epstein S, *et al*. Long-term proton pump inhibitor therapy and risk of Hip fracture. *JAMA* 2006;296:2947–53.
39. Coussens LM, Werb Z. Inflammation and cancer. *Nature* 2002;420:860–7.
40. González-Pérez A, Garcia Rodriguez L, Lopez-Ridaura R. Effects of non-steroidal anti-inflammatory drugs on cancer sites other than the colon and rectum: a meta-analysis. *BMC Cancer* 2003;3:28.
41. Mocellin S. Vitamin D and cancer: deciphering the truth. *Biochim Biophys Acta* 2011;1816:172–8.
42. Thomas E, Brewster DH, Black RJ, *et al*. Risk of malignancy among patients with rheumatic conditions. *Int J Cancer* 2000;88:497–502.
43. Bangalore S, Kumar S, Kjeldsen SE, *et al*. Antihypertensive drugs and risk of cancer: network meta-analyses and trial sequential analyses of 324,168 participants from randomised trials. *Lancet Oncol* 2011;12:65–82.
44. Vincent A, Herman J, Schulick R, *et al*. Pancreatic cancer. *Lancet* 2011;378:607–20.
45. Burbos N, Musonda P, Giarenis I, *et al*. Predicting the risk of endometrial cancer in postmenopausal women presenting with vaginal bleeding: the Norwich DEFAB risk assessment tool. *Br J Cancer* 2010;102:1201–6.
46. Stattin P, Björ O, Ferrari P, *et al*. Prospective study of hyperglycemia and cancer risk. *Diabetes Care* 2007;30:561–7.
47. Rustgi AK. *Cancers of the Gastrointestinal Tract. DeVita, Hellman, and Rosenberg's cancer: Principles & Practice of Oncology*. Wolters Kluwer/Lippincott Williams & Wilkins, Philadelphia, 2008:989–1313.
48. Perazella MA, Markowitz GS. Bisphosphonate nephrotoxicity. *Kidney Int* 2008;74:1385–93.
49. Hartmann LC, Sellers TA, Frost MH, *et al*. Benign breast disease and the risk of breast cancer. *N Engl J Med* 2005;353:229–37.
50. Mai PL, Garceau AO, Graubard BI, *et al*. Confirmation of family cancer history reported in a population-based survey. *J Natl Cancer Inst* 2011;103:788–97.
51. Chang ET, Smedby KE, Hjalgrim H, *et al*. Reliability of self-reported family history of cancer in a large case–control study of lymphoma. *J Natl Cancer Inst* 2006;98:61–8.
52. Heiss G, Wallace R, Anderson GL, *et al*. Health risks and benefits 3 years after stopping randomized treatment with estrogen and progestin. *JAMA* 2008;299:1036–45.
53. Yager JD, Davidson NE. Estrogen carcinogenesis in breast cancer. *N Engl J Med* 2006;354:270–82.
54. García Rodríguez LA, Lagergren J, Lindblad M. Gastric acid suppression and risk of oesophageal and gastric adenocarcinoma: a nested case control study in the UK. *Gut* 2006;55:1538–44.

Paper 5

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of gastrointestinal cancers: series of nested case-control studies with QResearch and CPRD data. *BMJ*. 2013;346:f114.

BMJ

BMJ 2013;356:f114 doi: 10.1136/bmj.f114 (Published 16 January 2013)

RESEARCH

Exposure to bisphosphonates and risk of gastrointestinal cancers: series of nested case-control studies with QResearch and CPRD data

Yana Vinogradova *research fellow in medical statistics*, Carol Coupland *associate professor and reader in medical statistics*, Julia Hippisley-Cox *professor of clinical epidemiology and general practice*

Division of Primary Care, University Park, Nottingham NG2 7RD, UK

Abstract

Objective To investigate the association between use of bisphosphonates estimated from prescription information and risk of gastrointestinal cancers.

Design Series of nested case-control studies.

Setting General practices in the United Kingdom contributing to the QResearch primary care database (660) and the Clinical Practice Research Datalink (CPRD) (643).

Participants Patients aged ≥50 with a diagnosis of a primary gastrointestinal cancer in 1997-2011, each matched with up to five controls by age, sex, practice, and calendar year.

Main outcome measures Odds ratios for incident gastrointestinal cancers (colorectal, oesophageal, gastric) and use of bisphosphonates, adjusted for smoking status, ethnicity, comorbidities, and use of other drugs.

Results 20 106 and 19 035 cases of colorectal cancer cases, 5364 and 5135 cases of oesophageal cancer cases, and 3155 and 3157 cases of gastric cancer were identified from QResearch and CPRD, respectively. Overall bisphosphonate use (at least one prescription) was not associated with risk of colorectal, oesophageal, or gastric cancers in either database. Adjusted odds ratios (95% confidence interval) for QResearch and CPRD were 0.97 (0.79 to 1.18) and 1.18 (0.97 to 1.43) for oesophageal cancer; 1.12 (0.87 to 1.44) and 0.79 (0.62 to 1.01) for gastric cancer; and 1.03 (0.94 to 1.14) and 1.10 (1.00 to 1.22) for colorectal cancer. Additional analyses showed no difference between types of bisphosphonate for risk of oesophageal and colorectal cancers. For gastric cancer, alendronate use was associated with an increased risk (1.47, 1.11 to 1.95; P=0.008), but only in data from the QResearch database and without any association with duration and with no definitive confirmation from sensitivity analysis.

Conclusions In this series of population based case-control studies in two large primary care databases, exposure to bisphosphonates was not associated with an increased risk of common gastrointestinal cancers.

Introduction

As an established drug for the treatment and prevention of osteoporosis,[1 2] bisphosphonates have been widely prescribed[3] and have a long term effect.[4] Although preclinical studies have shown that bisphosphonates have anti-tumour properties,[5 6] there is still a possibility that their adverse effects on the gastrointestinal tract, such as mucosal irritation, might cause ulceration[7] and could be linked to an increased risk of cancer.

The first publication on the possible association was from the US Food and Drug Administration (FDA) Adverse Event Reporting System, which listed 23 cases of oesophageal cancer in users of oral alendronate between 1995 and 2008.[8] An observational study, however, showed a reduced risk for oesophageal cancer but not gastric cancer.[9] A nested case-control study, based on the General Practice Research Database (GPRD), showed a 30% increased risk of oesophageal cancer in bisphosphonate users,[10] rising to more than a twofold increase in risk for more than three years' use, but it did not find a significant association with risk of gastric or colorectal cancers. A cohort study based on the GPRD, however, did not find any significant association between bisphosphonate use and risk of gastric or oesophageal cancers.[11 12] One Danish cohort study looked at gastrointestinal cancers and reported an increased risk of oesophageal cancer associated with use of alendronate and a possible protective effect of higher doses for colorectal cancer.[13] Finally, another Danish cohort study showed a reduced risk of gastric cancer and no excess risk in oesophageal cancer[14] in alendronate users. As for colorectal cancer, another cohort analysis that used the GPRD found a reduced risk associated with bisphosphonate use,[11] and an Israeli study also showed a significantly decreased risk in patients taking bisphosphonates for more an year.[15] Although a Danish study on postmenopausal women showed a reduced risk of colorectal cancer with oral bisphosphonates, the association was not time or dose dependent.[16]

Correspondence to: Y Vinogradova yana.vinogradova@nottingham.ac.uk

In summary, studies to date have reported conflicting findings, were based on data collected only up to 2008, and were limited by statistical power. We therefore investigated the association between bisphosphonates used for the prevention or treatment of osteoporosis and the risk of gastrointestinal cancers in the general population with a nested case-control design and including the most recent data from the QResearch database in the United Kingdom. We also replicated the analyses using the Clinical Practice Research Datalink (CPRD, previously known as General Practice Research Database (GPRD)).

Methods

Study design

The protocol for this study was published in 2012[17] and identified the QResearch UK primary care database as a source of data. Simultaneously, with the same protocol, a replicate study was conducted with CPRD. These databases are the largest primary care datasets in the UK and contain electronic records from 660 (QResearch) and 643 (CPRD) general practices, which include patients' demographics, referrals, tests, and prescriptions. Both have been successfully validated with other sources of information[18 19] and have been used for a range of safety studies involving commonly prescribed drugs.[10 20 21 22]

We identified open cohorts of patients aged ≥50 and registered with the practice at some time during the study period (January 1997 to July 2011). For this paper we selected gastrointestinal cancers (oesophageal, gastric, and colorectal) as the outcome and identified incident cases from the cohorts. The design of the study was a nested case-control study. We excluded patients aged <50 because the risk of the cancers of interest is low in this group and because bisphosphonates are rarely prescribed in younger people. We matched each case with up to five controls by age, sex, practice, and calendar year. All controls were alive and registered with the practice at the date of the first recorded diagnosis of cancer in their matched case, which we defined as the index date for each case and their matched controls. We excluded cases and controls with prescriptions for bisphosphonates licensed for any malignancies before the index date. Patients with Paget's disease were also excluded as their treatment requires higher doses of bisphosphonates and for much longer periods. Patients were included only if they had at least two years of data before their index date to ensure the completeness of records.

Exposure to bisphosphonates

The primary measure of exposure to bisphosphonates was assessed from prescription information within the observation period from the date of patient's registration with the practice to six months before the index date. Prescriptions in the last six months were ignored to reduce protopathic bias—early symptoms of cancer might include weight loss and low bone mineral density and be mistaken for symptoms of osteoporosis. A patient was considered to be a bisphosphonate user if they had at least one prescription in the observation period. The bisphosphonates to be included were identified in the *British National Formulary* section 6.6.2 as treatment for osteoporosis: alendronate, etidronate, ibandronate, and risedronate. We considered type of regimen for bisphosphonate use—daily or weekly/monthly—because daily use has been shown to have lower adherence than weekly use, with one of the possible reasons being inconvenience,[23] and we investigated possible differences between the effects of daily and other exposures to bisphosphonates on risk of gastrointestinal cancers. Cumulative exposure was estimated by summing the prescribed durations of prescription for each patient, considering gaps of fewer than 60 days between proximate prescriptions as continuous treatment. For drugs prescribed in cycles, the cycle length was taken to be the duration of prescription. Duration of use of bisphosphonates in the observation period (excluding the six months before the index date) was analysed with the following categorisations: no use; <6 months; 7-36 months; 37-72 months; ≥73 months. A test for trend was performed with the actual number of months.

Confounding variables

All the analyses included potential confounders established as risk factors for cancer. Body mass index (BMI)[24] as a continuous variable was measured at the date closest to one year before the index date. Smoking status[25] (current smoker: light (1-9 cigarettes/day), medium (10-19), heavy (≥20), ex-smoker, non-smoker); alcohol consumption[26] using Read codes for alcohol status; and ethnicity[27] (white or not recorded, black, Asian, other) were based on the latest values recorded before the index date. The analysis also adjusted for history of osteoporosis,[28] including a diagnosis of osteoporosis or osteopenia or previous fractures recorded before the index date; use of drugs that impair calcium absorption and reduce bone density or that affect risks of cancer such as systemic corticosteroids and acid suppressive drugs (including H_2 antagonists (*BNF* 1.3.1), proton pump inhibitors (*BNF* 1.3.5), and antacids (*BNF* 1.1.1))[29 30]; use of anti-inflammatory drugs[31] (traditional non-steroidal (NSAIDs), cyclo-oxygenase-2 inhibitors and aspirin)[32]; and use of vitamin D[33] if they were prescribed at least one year before the index date. Use of acid suppression drugs was categorised by the number of prescriptions within the observation period: none; <12 prescriptions; 12-24 prescriptions; 25-48 prescriptions; and ≥49 prescriptions to distinguish between occasional and more regular use.

We also included comorbidities that affect the risks of cancer (such as rheumatoid arthritis[34] for any cancer and diabetes for colorectal cancer[35]) if they were diagnosed at least a year before the index date. Analyses were adjusted for gastrointestinal disorders[36] if they were diagnosed before the first use of bisphosphonates or 12 months before the index date, whichever was earlier; diseases included upper gastrointestinal disease (dysphagia, oesophagitis, gastro-oesophageal reflux disease, hiatus hernia, oesophageal ulcers, Barrett's oesophagus, gastritis, duodenitis, peptic ulcers, dyspepsia), Crohn's disease, and ulcerative colitis. The results were also adjusted for family history of cancer[37] (this varied according to the cancer under consideration) if it was recorded at least six months before the index date. This is to reduce recall bias as patients are more likely to report a family history of cancer around the time of diagnosis.[38]

Statistical analysis

Analyses for these two studies were carried out separately in the two databases. We used conditional logistic regression to estimate odds ratios with 95% confidence intervals for cancer for each selected site and Wald test to examine the effects of duration and the differences in these between types of bisphosphonates. Missing values for the confounding factors (BMI, smoking status, and alcohol intake) were imputed with multiple imputation with the ICE programs in Stata.[39 40] We created 10 imputed datasets including all potential confounders and bisphosphonate exposure in the models and combined the results using Rubin's rules.[39] Results from both analyses were

pooled with the Mantel-Haenszel method for fixed effect models.

The primary analyses were based on exposure to bisphosphonates excluding prescriptions in the six months before the index date. We were carried out five sensitivity analyses. Firstly, having only one prescription in their records might mean that a patient never started bisphosphonate treatment or stopped it early because of adverse side effects, so we ran an analysis defining use of bisphosphonates as at least two prescriptions, but still excluding any in the last six months before the index date. In the second sensitivity analysis we included prescriptions in the last six months and considered a patient to be a user if they had at least one prescription in their records at any time before the index date.

Another possible bias in the main analysis might arise from the different observation times for patients so we undertook a third sensitivity analysis selecting patients only if they had at least six years of records and including prescriptions only between 6 and 72 months before the index date. Because Townsend score as a measure of deprivation was available for only 49% of CPRD practices we did not include it as a confounding variable in the main analyses in either database, so a fourth sensitivity analysis, restricted to patients with a valid Townsend score and adjusting for this, was run in both databases. A fifth sensitivity analysis was run with data only for patients with complete records for BMI, smoking, and alcohol intake. For the fourth and fifth sensitivity analyses, the definitions for use of bisphosphonates and years of medical records were identical to those in the main analysis.

All available data were used in the analyses and to allow for multiple comparisons we considered P<0.01 as significant, but to create a parity of presentation with other studies we have quoted a 95% confidence interval in our results. Sample size calculations are presented in the protocol.[17] Stata version 12 was used for the analyses.

Results

Study population

Within the study period, in QResearch we identified 20 106 cases of colorectal cancer, 5364 cases of oesophageal cancer, and 3155 cases of gastric cancer matched to 93 954, 25 101, and 14 715 controls, respectively. From CPRD, there were 19 035 cases of colorectal cancer, 5132 cases of oesophageal cancer, and 3157 cases of gastric cancer matched to 89 111, 24 053, and 14 686 controls, respectively.

Tables 1, 2 and 3⇓⇓⇓ show the characteristics for all cases and controls for both databases. Most of the descriptive statistics were similar in QResearch and CPRD. Cases and controls from QResearch were slightly younger than CPRD cases and controls, and Townsend score as a measure of deprivation was available only for a third of cases and controls in CPRD (table 1⇓). BMI, smoking status, and alcohol consumption had slightly fewer missing values in CPRD with higher proportions of non-smokers and moderate and high alcohol consumption than in QResearch (table 2⇓).

Most of the comorbidities had similar proportions in cases selected from both databases and their controls (table 3⇓). Upper gastrointestinal morbidities were recorded slightly more often in CPRD (29% in cases and 27% in controls *v* 22% and 20% in QResearch) with higher proportions for gastro-oesophageal reflux (9% and 8% *v* 6% and 5%, respectively) and dyspepsia (13% and 12% *v* 7% and 7%, respectively). Use of common drug treatments was similar in the databases with slightly more frequent prescribing of NSAIDs and less frequent prescribing of corticosteroids in CPRD compared with QResearch. Prescribing of calcium supplements also had different patterns with more patients prescribed calcium in CPRD.

Patterns of bisphosphonate use

In QResearch 4.6% of cases and 4.5% of controls had one or more prescriptions for bisphosphonates, as did 4.8% and 4.6%, respectively, in CPRD. About two thirds of patients with a diagnosis of osteoporosis had been prescribed bisphosphonates (64% of cases and 65% of controls in QResearch and 61% and 60%, respectively, in CPRD) and 2% of cases and controls in both databases had prescriptions for bisphosphonates without records of osteoporosis. Bisphosphonate users were more likely to be women and to have a lower BMI. Upper gastrointestinal problems were slightly more common (QResearch: cases 25% in users *v* 22% in non-users, controls 25% *v* 20%; CPRD: cases 35% *v* 29%, controls 33% *v* 27%) and use of acid lowering drugs was much more common in bisphosphonate users (QResearch: cases 62% in users *v* 37% in non-users, controls 61% *v* 32; CPRD: cases 65% *v* 39%, controls 63% *v*35%). Among bisphosphonate users, the proportion of patients with rheumatoid arthritis was more than five times higher than in non-users, and there were similar patterns for use of anti-inflammatory drugs, non-steroidal drugs, and corticosteroids.

More than three quarters of the bisphosphonate users were prescribed only one type of drug (77% in cases and 77% in controls in QResearch and 80% and 79% in CPRD), one fifth had prescriptions for two different types (20% in cases and 20% in controls in QResearch and 17% and 19% in CPRD), and less than 3% (2% in cases and 3% controls in QResearch and 3% and 2% in CPRD) had prescriptions for three different types during the observation period. Alendronate was the most commonly prescribed type (69% in cases and 66% in control users in QResearch and 69% and 69% in CPRD) and mostly prescribed for weekly use (87% in cases, 88% in controls in QResearch and 89% and 88% in CPRD). The second most common type was etidronate (34% in cases and 36% in controls users in QResearch and 30% and 32% in CPRD) with daily use for 14 days in 90 day cycles. The third most common was risedronate (21% in cases and 22% controls users in QResearch and 22% and 21% in CPRD), prescribed mostly for weekly use (76% in cases and 80% in controls users in QResearch and 82% and 81% in CPRD). Ibandronate (2% in cases and in controls users in QResearch and CPRD) was prescribed for monthly use only and no one received it as injections. In only one case (in QResearch) was a patient prescribed zoledronic acid.

In both databases, the minimum duration of bisphosphonate prescription was one week, and over half of bisphosphonate users had prescriptions for at least 20 months (median 20 (interquartile range 7-43) and 21 (8-44) for cases and controls, respectively, for QResearch; 19 (6-4) and 20 (7-41) for CPRD). Two thirds of bisphosphonate users (64% cases and 65% controls in QResearch, and 63% cases and controls in CPRD) had no gap of longer than 60 days between the first and last prescriptions, with 17% of cases and 18% of controls in QResearch and 19% of cases and 20% of controls in CPRD having only one gap longer than 60 days.

Associations with cancer

Tables 4⇓, 5⇓, and 6⇓ show the associations between regimen and duration of bisphosphonate prescriptions and different types of the drug and risk for oesophageal, gastric, and colorectal

cancer. Table 7⇓ contains the results from the first three sensitivity analyses for short and long term use for the three cancers.

Oesophageal cancer

After adjustment for confounders, both studies showed no significant association between overall bisphosphonate use and risk of oesophageal cancer (adjusted odds ratio 0.97, 95% confidence interval 0.79 to 1.18, for QResearch; and 1.18, 0.97 to 1.43, for CPRD). Similarly, there were no differences for frequency of use or duration (P=1.0 for trend) in QResearch. In CPRD, although odds ratios were progressively higher for longer use of bisphosphonates, none of them nor the trend test reached significance (P=0.07 for trend). There were no significant associations for individual types of bisphosphonate. None of the sensitivity analyses showed any significant associations.

Gastric cancer

After adjustment for confounders, both studies showed no significant association between overall bisphosphonate use and risk of gastric cancer (adjusted odds ratio 1.12, 95% confidence interval 0.87 to 1.44, for QResearch, and 0.79, 0.62 to 1.01, for CPRD). Daily use of bisphosphonates was associated with a decreased risk (0.60, 0.41 to 0.87; P=0.008) in CPRD, but this was not confirmed by any of the sensitivity analyses. Of the different bisphosphonates, alendronate use was significantly associated with cancer risk only in QResearch (1.47, 1.11 to 1.95; P=0.008), but a direct test between the different types of bisphosphonates was not significant (P=0.053). Short term use (<1 year) of alendronate was associated with a significantly higher risk of cancer in QResearch (1.91, 1.34 to 2.72; P<0.001), but there was no significant increase for longer term use (1.08, 0.74 to 1.59; P=0.7). These findings were also significant in the first sensitivity analysis, which classified use as two or more prescriptions (2.23, 1.54 to 3.22; P<0.001 for shorter term), but results of the second sensitivity analysis, which classified use as all prescriptions including the last six months, failed to reach significance (1.49, 1.06 to 2.08; P=0.02 for shorter term). In CPRD, the odds ratio for alendronate use (0.93, 0.71 to 1.22) was also higher than for other bisphosphonates, but the difference between drugs was not significant (P=0.2). There were no significant relations with duration for any of the drugs. No other significant associations were observed in the sensitivity analyses.

Colorectal cancer

Use of bisphosphonates was similar in cases of colorectal cancer and their matched controls (4.6% in both), and adjustment for the confounders did not show any association between bisphosphonate use and risk of cancer (adjusted odds ratio 1.03, 95% confidence interval 0.94 to 1.14, for QResearch; and 1.10, 1.00 to 1.22, for CPRD). Daily use had similar effects as weekly or monthly use. In CPRD data, short term use was associated with an increased risk, although this reached significance only for one to six months' use (1.27, 1.09 to 1.48; P=0.002), but this finding was not confirmed by any of the sensitivity analyses. The relation between duration of bisphosphonate use and risk of colorectal cancer was not significant in either database (P=0.3 for trend for QResearch and P=0.5 for CPRD). For QResearch, the adjusted odds ratio for any use of alendronate was higher than for etidronate or risedronate, but it was not significantly increased and the difference between the drugs was not significant (P=0.09). For CPRD, the effect of drugs seemed to be similar (P=0.5). The results from the sensitivity analyses were in line with these.

Results obtained from pooling the results from both databases presented in tables 4, 5, and 6 did not show any significant findings.

Discussion

Summary

This series of case-control studies on two large population databases found no overall association between use of bisphosphonates and risk of oesophageal, gastric, or colorectal cancer. There was a small significantly increased risk of gastric cancer associated with use of alendronate in one database, which was restricted to short term alendronate users; the risk was nearly twofold for patients who used alendronate for less than a year. This is unlikely to be a causal relation as there was no association with longer term use. For colorectal and oesophageal cancers, there were no associations that suggested an increased risk of these cancers in people using bisphosphonates.

Strengths and limitations

This was the first drug safety study to be undertaken with both CPRD and QResearch databases with identical definitions for confounders, exposures, and sampling to help ensure comparability. As an observational study based on routinely collected data from two large primary care research databases, it has the strengths and limitations common to such studies. Our study was substantially larger and had much greater statistical power than any previous study. This allowed analyses to be carried out to investigate the effects of duration of treatment on risk of cancer. As the data on prescriptions and potential confounding variables were routinely and prospectively collected and recorded before the index date, the study was free from recall bias. Because all eligible cases and randomly selected controls were included, there was also no selection bias.

The limitations of the study include possible uncertainty in records of diagnosis of cancer. A systematic review based on GPRD validation studies reported that, on average, 95% of diagnoses of cancer recorded on the general practice electronic record were confirmed from other data sources.[41] It has been found, however, that one in five of all primary care patients with cancer were not identified through electronic searches for malignancies in general practice electronic records, although this was based on data from 1990-99.[42] Any such misclassification might result in underestimation of associations with bisphosphonates, shifting odds ratios toward unity. Also, the selection of cases was based on the first record of a cancer while the exact origin site might have been determined only later, and this level of detail was not available across all records. Information about cancer stage or results of histological investigations was also not consistently recorded in general practice, so was not used. Within each site of origin, this creates an inability to distinguish between specific cancers that might have different risk factors.

Another limitation is that there might have been an overestimation of bisphosphonate use. The analyses were based on prescriptions rather than actual use, and no data were available on adherence to treatment. There is, however, no reason for non-adherence to systematically differ between cases and controls.

There could have been some residual confounding as information on some risk factors such as bone mineral density, physical activity, diet, and cancer screening tests (endoscopy

or colonoscopy) is not generally recorded so these factors were not included in the analyses.

Bisphosphonate users

Comparisons between bisphosphonate users and non-users reflected the recommendations for targeting groups with osteoporosis, as some characteristics and comorbidities, such as low BMI and rheumatoid arthritis in users, are associated with diagnosis of primary osteoporosis. A heightened rate of upper gastrointestinal disorders before onset of treatment for osteoporosis has also been reported elsewhere.[43] As secondary osteoporosis is more likely to develop in patients taking acid lowering drugs[29] and corticosteroids,[44 45] the proportion of such patients was noticeably higher in bisphosphonate users. Although upper gastrointestinal disorders and use of acid lowering drugs could be important confounders, they might also lie on the causal pathway. Removal of them from the adjusted analyses, however, did not noticeably change the results.

Oesophageal cancer

Oesophageal cancer associated with bisphosphonate use has been of most concern to epidemiologists. So far, however, the studies of this association have all been much smaller than our study, which was based on 252 bisphosphonate users with oesophageal cancer in QResearch and 262 in CPRD, and have yielded inconsistent results.

The study by Green and colleagues, which was based on an earlier version of CPRD with larger proportions of missing values and more exposure to etidronate than alendronate (57 *v* 37 exposed cases), reported an association with duration of treatment, with an over twofold increase in people who had been taking these drugs for more than three years (33 such cases).[10] These results were, however, adjusted only for smoking status, alcohol intake, and BMI, without adjustment for the important confounders of osteoporosis and use of any drugs other than bisphosphonates associated with its treatment, so possibly overestimating cancer risk. Similarly, another recent study on GPRD data from 1995 to 2007 showed an increased risk of oesophageal cancer associated with bisphosphonate use, but only in women (adjusted odds ratio 1.43, 95% confidence interval 1.16 to 1.75), and again the results were not adjusted for osteoporosis or drugs associated with it.[46] There seemed to be an opposite association for men (adjusted odds ratio 0.73, 0.53 to 1.03), although it was not significant. A Taiwanese case-control study analysed prescription information from 16 204 cases and 64 816 controls and reported no association for overall exposure and increased risk only in rare users, but this study did not adjust for any confounding factors.[47]

The cohort study by Cardwell and colleagues,[11] which also used the CPRD database, identified 79 cases of oesophageal cancer in bisphosphonate users and did not report a significantly increased risk of cancer even for cumulative exposure of more than three years (hazard ratio 0.90, 0.44 to 1.81). In our CPRD analyses, use for more than three years was associated with an increased risk (92 such cases, odds ratio 1.38, 95% confidence interval 1.04 to 1.84; P=0.03), but this was not significant at the 1% level. In our QResearch analyses, we did not observe an increased risk of oesophageal cancer in bisphosphonate users but rather showed a reduced, but again not significant, risk for people who had been taking these drugs for more than three years (75 such cases, odds ratio 0.92, 0.68 to 1.25).

Although Vestergaard's study showed significantly increased risks of oesophageal cancer in etidronate and alendronate users (32 and 14 cases, respectively),[13] these associations were not time dependent and therefore might not be causal. Another cohort study looking at alendronate and etidronate use reported a decreased risk of oesophageal cancer for both,[14] although not significant, and showed even greater reduction in patients with 10 prescriptions or more, but this was still not significant.

Gastric cancer

In line with our findings, no association between use of any bisphosphonates and gastric cancer has been reported in previous CPRD publications.[10 12] Our CPRD analyses had similar findings to those of Green and colleagues[10] for gastric cancer on the same database and, despite a bigger sample, did not reach a significant level for the association of a decreased risk with bisphosphonate use. Our QResearch analyses showed an opposite association for gastric cancer and use of alendronate to the one found in the study by Abrahamsen and colleagues (39% reduced risk based on 22 exposed cases)[14] but our association, an almost twofold increased risk, was seen only in short term users, with no association for long term use. In our CPRD analyses, however, we found no association with alendronate use. It is possible that this short term association resulted from users stopping bisphosphonate treatment because the drugs caused side effects similar to symptoms of gastric cancer or aggravated early symptoms of gastric cancer, leading to earlier detection of existing cancer. It has been shown that alendronate in particular might be associated with gastrointestinal mucosal injury,[48] which would require additional tests. Given the lack of consistency between the CPRD and QResearch results, however, this might simply be a spurious chance finding, and it is unlikely to reflect a causal association.

The paper by Cardwell and colleagues noted that there are possible misclassifications of gastric and oesophageal cancers, and that study showed no association between bisphosphonate use and risk of those cancers when combined.[11] To check this, we ran a combined analysis to compare our findings and also did not observe any significant association in either database (overall use for QResearch odds ratio 1.02, 95% confidence interval 0.88 to 1.20; and 1.02, 0.87 to 1.18, for CPRD), all P values being greater than 0.1 except for short term alendronate use in QResearch (odds ratio 1.29, 1.03 to 1.60; P=0.02).

Colorectal cancer

Based on 20 106 cases of colorectal cancer in QResearch and 19 035 in CPRD with, respectively, 929 and 902 cases exposed to bisphosphonates, our study found no association between bisphosphonate use and risk of colorectal cancer. The increased risk found in the CPRD analysis and associated only with short term use (up to six months) was inconsistent with the QResearch findings and not confirmed in sensitivity analyses. Similarly inconsistent results have been found in other smaller epidemiological studies.[10 12 13] Two studies—Cardwell et al[12] and Green et al[10]—that used CPRD and based, respectively, on 264 and 276 bisphosphonate users, found similar associations for risk of colorectal cancer, but reached significant levels only in one[12] (hazard ratio for any use 0.74, 95% confidence interval 0.60 to 0.91). Cardwell and colleague's study,[12] however, used an earlier version of the same data source and had much higher proportions of missing data for BMI, smoking, and alcohol consumption, all factors associated with both bisphosphonate prescribing and risk of cancer. Another study, which used an Israeli health services database, also reported a decreased risk of colorectal cancer in bisphosphonate users (odds ratio for any use 0.67, 95% confidence interval 0.51 to 0.88).[15] This was based on only 97 cases in users and was subject to recall and selection biases, with different response rates for cases and

controls (83% and 58%, respectively). An American study on menopausal women enrolled in the Nurses Health Study reported no association either for overall use of bisphosphonates or for different terms of use.[49] Although a recent Danish study showed an association between overall use of bisphosphonate and reduced risk of colorectal cancer (hazard ratio 0.69, 95% confidence interval 0.60 to 0.79) and even described a possible mechanism based on anti-tumour properties of the drug, the study did not show any dose or time dependent associations, and the risk of cancer was not significantly reduced in long term (more than six months) users (hazard ratio 0.78, 0.55 to 1.11).[16]

Conclusion

We have conducted a series of large population based case-control studies using the two largest primary care databases in the UK to examine the association of bisphosphonates with risks of common gastrointestinal cancers in the general population. The databases provided almost the same numbers of cases and controls and the recorded data were similar in demographics, lifestyle related factors, comorbidities, and use of drugs. Although several findings were consistent between the databases, we found no increased risk in general for any of the cancers except for an increased risk of gastric cancer in alendronate users in one of the databases, but without time associations.

We acknowledge the contribution of EMIS and the University of Nottingham for expertise in creating and maintaining QResearch, and to the EMIS practices that contribute data, without whom this research would not be possible.

Contributors: JH-C had the original idea for this study. CC contributed to the development of the idea and the study design. YV reviewed the literature, contributed to the study design, undertook the primary analysis and the first interpretation, and wrote the first draft of the paper. JH-C and CC critically reviewed the paper. All authors approved the submitted version. YV is guarantor.

Funding: This work was funded by the division of primary care of University of Nottingham.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no support from any additional organisation for the submitted work. JH-C is an unpaid director of QResearch, a not-for-profit organisation that is a joint partnership between the University of Nottingham and EMIS (commercial IT supplier for 60% of general practices in the UK). JH-C is also a paid director of ClinRisk, which produces open and closed source software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help improve patient care.

Ethical approval: This protocol has been published in BMJ Open. It has also independently peer reviewed by the QResearch Scientific Board and has been reported to Trent research ethics committee in accordance with the agreed procedure (reference No MREC/03/4/021). For CPRD data analysis, the protocol was approved by independent scientific advisory committee (reference No ISAC 11_149).

Data sharing: Descriptive statistics for users and non-users of bisphosphonates and results of unadjusted and sensitivity analyses are available from the corresponding author.

1 National Institute for Health and Clinical Excellence. Alendronate, etidronate, risedronate, strontium ranelate and raloxifen for preventing bone fractures in postmenopausal women with osteoporosis who have not had a fracture. Information about NICE technology appraisal guidance No 160. NICE, 2008.
2 National Institute for Health and Clinical Excellence. Alendronate, etidronate, risedronate, strontium ranelate and teriparatide for preventing bone fractures in postmenopausal women with osteoporosis who have already had a fracture. Information about NICE technology appraisal guidance No 161. NICE, 2008.
3 Kanis J, McCloskey E, Johansson H, Strom O, Borgstrom F, Oden A. Case finding for the management of osteoporosis with FRAX®—assessment and intervention thresholds for the UK. *Osteoporos Int* 2008;19:1395-408.
4 Watts NB, Diab DL. Long-term use of bisphosphonates in osteoporosis. *J Clin Endocrinol Metab* 2010;95:1555-65.
5 Guise TA. Antitumor effects of bisphosphonates: promising preclinical evidence. *Cancer Treat Rev* 2008;34(suppl 1):S19-24.
6 Croucher P, Jagdev S, Coleman R. The anti-tumor potential of zoledronic acid. *Breast* 2003;12(suppl 2):S30-36.
7 Zografos GN, Georgiadou D, Thomas D, Kaltsas G, Digalakis M. Drug-induced esophagitis. *Dis Esophagus* 2009;22:633-37.
8 Wysowski DK. Reports of esophageal cancer with oral bisphosphonate use. *N Engl J Med* 2009;360:89-90.
9 Abrahamsen B, Eiken P, Eastell R. More on reports of esophageal cancer with oral bisphosphonate use. *N Engl J Med* 2009;360:1791-92.
10 Green J, Czanner G, Reeves G, Watson J, Wise L, Beral V. Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort. *BMJ* 2010;341:c4444.
11 Cardwell CR, Abnet CC, Cantwell MM, Murray LJ. Exposure to oral bisphosphonates and risk of esophageal cancer. *JAMA* 2010;304:657-63.
12 Cardwell CR, Abnet CC, Veal P, Hughes CM, Cantwell MM, Murray LJ. Exposure to oral bisphosphonates and risk of cancer. *Int J Cancer* 2012;131:E717-25.
13 Vestergaard P. Occurrence of gastrointestinal cancer in users of bisphosphonates and other antiresorptive drugs against osteoporosis. *Calcif Tissue Int* 2011;89:434-41.
14 Abrahamsen B, Pazianas M, Eiken P, Russell RGG, Eastell R. Esophageal and gastric cancer incidence and mortality in alendronate users. *J Bone Mineral Res* 2011;27:679-86.
15 Rennert G, Pinchev M, Rennert HS, Gruber SB. Use of bisphosphonates and reduced risk of colorectal cancer. *J Clin Oncol* 2011;29:1146-50.
16 Pazianas M, Abrahamsen B, Eiken P, Eastell R, Russell R. Reduced colon cancer incidence and mortality in postmenopausal women treated with an oral bisphosphonate—Danish National Register Based Cohort Study. *Osteoporos Int* :2012;23:2693-701.
17 Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of cancer: a protocol for nested case-control studies using the QResearch primary care database. *BMJ Open* 2012;2:e000548.
18 Hippisley-Cox J, Hammersley V, Pringle M, Coupland C, Crown N, Wright L. How useful are general practice databases for research? Analysis of their accuracy and completeness in one research network. *Health Inform J* 2004;10:91-109.
19 Jick SS, Kaye JA, Vasilakis-Scaramozza C, Rodríguez LAG, Ruigómez A, Meier CR, et al. Validity of the general practice research database. *Pharmacotherapy* 2003;23:686-9.
20 Hippisley-Cox J, Coupland C. Unintended effects of statins in men and women in England and Wales: population based cohort study using the QResearch database *BMJ* 2010;340:c2197.
21 Parker C, Coupland C, Hippisley-Cox J. Antipsychotic drugs and risk of venous thromboembolism: nested case-control study. *BMJ* 2010;341:c4245.
22 Walker AJ, Card T, Bates TE, Muir K. Tricyclic antidepressants and the incidence of certain cancers: a study using the GPRD. *Br J Cancer* 2011;104:193-97.
23 Recker RR, Gallagher R, MacCosbe PE. Effect of dosing frequency on bisphosphonate medication adherence in a large longitudinal cohort of women. *Mayo Clin Proc* 2005;80:856-61.
24 Henderson KD, Bernstein L. Etiology of cancer: obesity and physical activity. In: DeVita, Hellman, and Rosenberg's Cancer: principles and practice of oncology. Wolters Kluwer/Lippincott Williams and Wilkins, 2008:239-44.
25 Hecht SS. Etiology of cancer: tobacco. In: DeVita, Hellman, and Rosenberg's Cancer: principles and practice of oncology. Wolters Kluwer/Lippincott Williams and Wilkins, 2008:147-55.
26 Schütze M, Boeing H, Pischon T, Rehm J, Kehoe T, Gmel G, et al. Alcohol attributable burden of incidence of cancer in eight European countries based on results from prospective cohort study. *BMJ* 2011;342:d1584.
27 Ferlay J, Shin H-R, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127:2893-917.
28 McGlynn KA, Gridley G, Mellemkjaer L, Brinton LA, Anderson KC, Caporaso NE, et al. Risks of cancer among a cohort of 23,935 men and women with osteoporosis. *Int J Cancer* 2008;122:1879-84.
29 Corley DA, Kubo A, Zhao W, Quesenberry C. Proton pump inhibitors and histamine-2 receptor antagonists are associated with hip fractures among at-risk patients. *Gastroenterology* 2010;139:93-101.
30 Rodríguez LAG, Lagergren J, Lindblad M. Gastric acid suppression and risk of oesophageal and gastric adenocarcinoma: a nested case control study in the UK. *Gut* 2006;55:1538-44.
31 Coussens LM, Werb Z. Inflammation and cancer. *Nature* 2002;420:860-67.
32 Gonzalez-Perez A, Garcia Rodriguez L, Lopez-Ridaura R. Effects of non-steroidal anti-inflammatory drugs on cancer sites other than the colon and rectum: a meta-analysis. *BMC Cancer* 2003;3:28.
33 Mocellin S. Vitamin D and cancer: deciphering the truth. *Biochim Biophys Acta* 2011;1816:172-8.
34 Thomas E, Brewster DH, Black RJ, Macfarlane GJ. Risk of malignancy among patients with rheumatic conditions. *Int J Cancer* 2000;88:497-502.
35 Stattin P, Björ O, Ferrari P, Lukanova A, Lenner P, Lindahl B, et al. Prospective study of hyperglycemia and cancer risk. *Diabetes Care* 2007;30:561-7.
36 Rustgi AK. Cancers of the gastrointestinal tract. DeVita, Hellman, and Rosenberg's Cancer: principles and practice of oncology Wolters Kluwer/Lippincott Williams and Wilkins, 2008:989-1313.
37 Mai PL, Garceau AO, Graubard BI, Dunn M, McNeel TS, Gonsalves L, et al. Confirmation of family cancer history reported in a population-based survey. *J Natl Cancer Inst* 2011;103:788-97.
38 Chang ET, Smedby KE, Hjalgrim H, Glimelius B, Adami H-O. Reliability of self-reported family history of cancer in a large case-control study of lymphoma. *J Natl Cancer Inst* 2006;98:61-8.
39 Royston P. Multiple imputation of missing values. *Stata J* 2004;4:227-41.
40 Royston P. Multiple imputation of missing values: update of ICE. *Stata J* 2005;5:527-36.
41 Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010;69:4-14.

What is already known on this topic

Bisphosphonates have become a common treatment for osteoporosis

Although preclinical studies have shown bisphosphonates have anti-tumour properties, epidemiological evidence concerning the associations between bisphosphonates and risk of gastrointestinal cancers has not been consistent

What this study adds

This series of nested case-control studies was conducted with two large general population primary care databases and found no overall association between use of bisphosphonates and risk of oesophageal, gastric, or colorectal cancers

An increased risk of gastric cancer associated with use of alendronate was found in one database, but with no evidence of a duration response

42 Pascoe S, Neal R, Heywood P, Allgar V, Miles J, Stefoski-Mikeljevic J. Identifying patients with a cancer diagnosis using general practice medical records and cancer registry data. *Fam Pract* 2008;25:215-20.
43 Vestergaard P, Schwartz K, Pinholt E, Rejnmark L, Mosekilde L. Gastric and esophagus events before and during treatment of osteoporosis. *Calcif Tissue Int* 2010;86:110-5.
44 Mauck KF, Clarke BL. Diagnosis, screening, prevention, and treatment of osteoporosis. *Mayo Clin Proc* 2006;81:662-72.
45 Hippisley-Cox J, Coupland C. Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. *BMJ* 2009;339:b4229.
46 Wright E, Schofield PT, Seed P, Molokhia M. Bisphosphonates and risk of upper gastrointestinal cancer—a case control study using the General Practice Research Database (GPRD). *PLoS ONE* 2012;7:e47616.
47 Ho Y-F, Lin J-T, Wu C-Y. Oral Bisphosphonates and risk of esophageal cancer: a dose-intensity analysis in a nationwide population. *Cancer Epidemiol Biomarkers Prev* 2012;21:993-5.
48 Thomson ABR, Marshall JK, Hunt RH, Provenza JM, Lanza FL, Royer MG, et al. 14 day endoscopy study comparing risedronate and alendronate in postmenopausal women stratified by Helicobacter pylori status. *J Rheumatol* 2002;29:1965-74.
49 Khalili H, Huang ES, Ogino S, Fuchs CS, AT C. A prospective study of bisphosphonate use and risk of colorectal cancer. *J Clin Oncol* 2012;30:3229-33.

Accepted: 18 December 2012

Cite this as: *BMJ* 2013;356:f114

Tables

Table 1| Baseline demographic characteristics in cases (all gastrointestinal cancers) and all matched controls by database (QResearch or CPRD). Values are percentages and numbers

	QResearch		CPRD	
	Cases (n=28 625)	Controls (n=133 770)	Cases (n=27 324)	Controls (n=27 850)
Sex:				
Male	59.9 (17 146)	59.9 (80 083)	58.1 (15 871)	58.1 (74 310)
Female	40.1 (11 479)	40.1 (53 687)	41.9 (11 453)	41.9 (53 540)
Age band (years):				
50-54	4.9 (1401)	4.8 (6455)	4.9 (1338)	4.9 (6247)
55-64	20.5 (5876)	20.5 (27 424)	20.1 (5502)	20.3 (25 954)
65-74	33.6 (9632)	34.0 (45 511)	31.7 (8668)	32.3 (41 286)
75-84	34.0 (9723)	34.2 (45 688)	32.1 (8765)	32.5 (41 491)
≥85	7.0 (1993)	6.5 (8692)	11.2 (3051)	10.1 (12872)
Ethnicity:				
White	23.1 (6600)	21.1 (28 228)	7.7 (2107)	7.1 (9069)
Not recorded*	76.0 (21 754)	77.8 (104 022)	88.9 (24 281)	89.6 (114 513)
Other	0.9 (271)	1.1 (1520)	3.4 (936)	3.3 (4268)
Asian	0.4 (111)	0.6 (741)	0.2 (45)	0.3 (414)
Black	0.4 (103)	0.4 (506)	0.1 (40)	0.1 (169)
Other	0.2 (57)	0.2 (273)	3.1 (851)	2.9 (3685)
Deprivation (5th of Townsend score):				
1 (most affluent)†	25.4 (7059)	26.1 (33 900)	28.0 (2737)	29.6 (13 542)
2*	21.4 (5944)	22.5 (29 169)	24.2 (2358)	24.0 (10 970)
3*	20.7 (5768)	20.2 (26 163)	20.1 (1966)	19.8 (9029)
4*	18.3 (5099)	17.8 (23 131)	16.4 (1605)	15.8 (7212)
5, most deprived†	14.2 (3966)	13.4 (17 376)	11.2 (1092)	10.8 (4922)
Not recorded	2.8 (789)	3.0 (4031)	64.3 (17 566)	64.3 (82 175)

*Assumed white for analyses.

†Proportion only within recorded data.

Table 2| Baseline characteristics in cases (all gastrointestinal cancers) and all matched controls by database (QResearch or CPRD). Values are percentages and numbers unless stated otherwise

	QResearch		CPRD	
	Cases (n=28 625)	Controls (n=133 770)	Cases (n=27 324)	Controls (n=127 850)
Body mass index (BMI):				
15-24*	35.0 (7729)	36.2 (35 826)	36.4 (8242)	37.7 (39 188)
25-29*	42.0 (9290)	42.1 (41 650)	41.4 (9363)	41.6 (43 319)
30-49*	23.0 (5088)	21.7 (21 424)	22.2 (5029)	20.7 (21 535)
Not recorded	22.8 (6518)	26.1 (34 870)	17.2 (4690)	18.6 (23 808)
Years since recorded†	2.6 (1.5-5.6)	2.7 (1.5-5.9)	2.6 (1.5-5.9)	2.8 (1.5-6.2)
Smoking status:				
Non-smoker*	43.1 (11230)	47.3 (55 708)	50.1 (12 904)	54.4 (63 608)
Ex-smoker*	39.4 (10 268)	37.2 (43 894)	33.5 (8621)	30.8 (35 979)
Current light*	6.1 (1591)	5.5 (6485)	3.8 (985)	3.6 (4200)
Current moderate*	7.5 (1958)	6.7 (7870)	8.7 (2233)	7.9 (9217)
Current heavy*	3.9 (1010)	3.3 (3919)	3.9 (997)	3.4 (3927)
Not recorded	9.0 (2568)	11.9 (15 894)	5.8 (1584)	8.5 (10 919)
Years since recorded†	0.9 (0.3-3.0)	1.1 (0.4-3.6)	1.0 (0.3-3.2)	1.2 (0.4-3.8)
Use of alcohol:				
No use*	30.5 (6925)	30.8 (31 788)	26.4 (6114)	27.0 (28 540)
Ex-user*	—	—	2.6 (601)	2.4 (2489)
Light*	57.8 (13140)	59.8 (61 764)	55.7 (12 924)	57.3 (60 633)
Moderate and more*	11.7 (2653)	9.4 (9731)	15.4 (3564)	13.4 (14 192)
Not recorded	20.6 (5907)	22.8 (30 487)	15.1 (4121)	17.2 (21 996)
Years since recorded†	2.1 (0.6-5.5)	2.4 (0.7-6.0)	3.0 (0.9-7.0)	3.3 (1.0-7.4)

*Proportion only within recorded data.

‡Median (interquartile range) time in years between recorded value and index date.

Table 3| **Medical history at baseline in cases (all gastrointestinal cancers) and all matched controls by database (QResearch or CPRD). Values are percentages and numbers.**

	QResearch		CPRD	
	Cases (n=28 625)	Controls (n=133 770)	Cases (n=27 324)	Controls (n=127 850)
Comorbidities				
Upper gastrointestinal	22.1 (6339)	20.1 (26 935)	29.3 (8016)	26.8 (34 247)
Oesophagitis	7.3 (2103)	6.3 (8475)	7.4 (2032)	6.1 (7846)
Oesophageal ulcer	0.3 (86)	0.1 (176)	0.4 (109)	0.2 (225)
Barrett's oesophagus	1.0 (299)	0.6 (746)	1.1 (312)	0.5 (691)
Dysphagia	1.4 (400)	1.1 (1440)	2.8 (755)	2.3 (2954)
Gastro-oesophageal reflux	5.6 (1593)	5.1 (6858)	8.8 (2397)	8.0 (10 223)
Hiatus hernia	4.7 (1352)	4.2 (5560)	6.1 (1673)	5.4 (6923)
Gastritis	2.9 (822)	2.7 (3630)	3.9 (1065)	3.5 (4417)
Peptic ulcer	6.5 (1863)	5.6 (7529)	8.0 (2179)	6.8 (8750)
Duodenitis	0.7 (199)	0.7 (910)	0.9 (245)	0.9 (1141)
Dyspepsia	7.1 (2029)	6.6 (8839)	12.9 (3525)	12.1 (15 456)
Diabetes	11.0 (3157)	9.2 (12 258)	11.3 (3093)	9.6 (12 324)
Crohn's disease	0.2 (60)	0.2 (321)	0.3 (70)	0.3 (360)
Ulcerative colitis	0.9 (255)	0.6 (836)	0.8 (214)	0.7 (864)
Chronic kidney disease	4.1 (1178)	3.9 (5277)	4.6 (1269)	4.5 (5726)
Rheumatoid arthritis	1.5 (422)	1.6 (2081)	1.7 (471)	1.7 (2233)
History of osteoporosis				
Osteoporosis/osteopenia	4.0 (1146)	4.1 (5501)	4.6 (1269)	4.7 (5985)
Osteoporotic fractures	3.9 (1128)	4.0 (5326)	3.0 (811)	2.9 (3699)
Family history of cancer				
Colorectal cancer	0.3 (73)	0.2 (238)	0.3 (85)	0.3 (360)
Gastrointestinal cancer	1.1 (315)	0.9 (1217)	0.3 (74)	0.2 (220)
Drugs used (excluding last 6 months)				
Acid lowering drugs	38.4 (10 980)	33.7 (45 048)	40.6 (11 099)	36.0 (45 968)
NSAIDs	53.8 (15 394)	54.1 (72 346)	58.6 (16 012)	59.3 (75 769)
Corticosteroids	15.1 (4330)	13.6 (18 129)	14.4 (3929)	13.1 (16 696)
Calcium	1.1 (311)	1.1 (1534)	7.5 (2057)	7.6 (9667)
Vitamin D	5.7 (1634)	5.6 (7490)	6.0 (1629)	6.1 (7736)
Bisphosphonates				
Any	4.6 (1322)	4.5 (5996)	4.8 (1303)	4.6 (5868)
Alendronate	3.2 (906)	3.0 (3976)	3.3 (894)	3.2 (4038)
Etidronate	1.6 (451)	1.6 (2160)	1.4 (391)	1.4 (1850)
Risedronate	1.0 (273)	1.0 (1301)	1.1 (287)	1.0 (1236)
Ibandronate	0.1 (24)	0.1 (134)	0.1 (24)	0.1 (103)
Other osteoporosis drugs				
Any	0.3 (94)	0.3 (451)	0.4 (116)	0.4 (506)
Raloxifen	0.2 (59)	0.2 (261)	0.2 (62)	0.2 (289)
Strontium	0.1 (26)	0.1 (137)	0.2 (46)	0.1 (172)
Calcitonin	0.0 (14)	0.1 (77)	0.0 (13)	0.1 (71)

NSAIDs=non-steroidal anti-inflammatory drugs.

Table 4| Bisphosphonate use in oesophageal cancer cases and controls, numbers and odds ratios (95% confidence intervals) compared with non-use by database

	QResearch			CPRD			Combined analysis	
	Cases/controls	Adjusted* odds ratio (95% CI)	P value	Cases/controls	Adjusted* odds ratio (95% CI)	P value	Pooled odds ratio (95% CI)	P value
Total	5364/25 101	—	—	5132/24 053	—	—	—	—
Overall use								
None	5112/24 030	—	—	4870/23 110	—	—	—	—
At least 1 prescription	252/1071	0.97 (0.79 to 1.18)	0.7	262/943	1.18 (0.97 to 1.43)	0.09	1.07 (0.93 to 1.23)	0.3
Regimen of use								
Daily	103/387	1.09 (0.84 to 1.40)	0.5	96/337	1.16 (0.89 to 1.51)	0.3	1.12 (0.93 to 1.35)	0.2
Weekly or monthly	149/684	0.88 (0.70 to 1.12)	0.3	166/606	1.20 (0.95 to 1.50)	0.1	1.03 (0.88 to 1.22)	0.7
Duration of use								
Short term (<1 year)	94/367	1.03 (0.80 to 1.34)	0.8	89/361	1.05 (0.81 to 1.37)	0.7	1.04 (0.87 to 1.25)	0.7
Long term (≥1 year)	158/704	0.92 (0.73 to 1.16)	0.5	173/582	1.28 (1.02 to 1.60)	0.03	1.09 (0.93 to 1.28)	0.3
1-6 months	55/226	1.00 (0.72 to 1.38)	1.0	58/222	1.12 (0.82 to 1.54)	0.5	1.06 (0.84 to 1.33)	0.6
7-36 months	122/512	0.97 (0.76 to 1.23)	0.8	112/428	1.10 (0.86 to 1.41)	0.4	1.03 (0.87 to 1.22)	0.7
37-72 months	49/234	0.90 (0.63 to 1.28)	0.6	63/211	1.31 (0.94 to 1.81)	0.1	1.10 (0.87 to 1.40)	0.4
>72 months	26/99	1.02 (0.63 to 1.64)	0.9	29/82	1.63 (1.03 to 2.59)	0.04	1.30 (0.93 to 1.81)	0.1
Trend for months of use	—	—	1.0	—	—	0.07	—	0.2
Alendronate								
Any use	163/721	0.91 (0.73 to 1.14)	0.4	167/650	1.03 (0.83 to 1.28)	0.8	0.97 (0.83 to 1.13)	0.7
<1 year	81/309	1.03 (0.78 to 1.37)	0.8	72/297	0.97 (0.73 to 1.30)	0.8	1.00 (0.82 to 1.23)	1.0
≥1 year	82/412	0.82 (0.62 to 1.08)	0.2	95/353	1.10 (0.84 to 1.44)	0.5	0.95 (0.78 to 1.16)	0.6
Trend for months of use	—	—	0.2	—	—	0.1	—	0.8
Etidronate								
Any use	99/363	1.17 (0.91 to 1.50)	0.2	89/306	1.11 (0.85 to 1.45)	0.4	1.14 (0.95 to 1.37)	0.2
<1 year	35/141	1.02 (0.69 to 1.51)	0.9	34/126	1.06 (0.71 to 1.58)	0.8	1.04 (0.79 to 1.37)	0.8
≥1 year	64/222	1.24 (0.91 to 1.69)	0.2	55/180	1.15 (0.82 to 1.60)	0.4	1.20 (0.96 to 1.50)	0.1
Trend for months of use	—	—	0.04	—	—	0.6	—	0.06
Risedronate								
Any use	50/236	0.82 (0.59 to 1.16)	0.3	60/207	1.14 (0.83 to 1.57)	0.4	0.98 (0.78 to 1.24)	0.9
<1 year	22/107	0.79 (0.48 to 1.29)	0.3	24/103	0.91 (0.57 to 1.45)	0.7	0.85 (0.60 to 1.19)	0.3
≥1 year	28/129	0.81 (0.52 to 1.26)	0.4	36/104	1.39 (0.93 to 2.09)	0.1	1.09 (0.80 to 1.46)	0.6
Trend for months of use	—	—	0.1	—	—	0.3	—	0.9

†Adjusted for BMI, smoking status, alcohol consumption, ethnicity, rheumatoid arthritis, osteoporosis and fractures, use of other osteoporosis drugs, vitamin D, NSAIDs, corticosteroids, acid lowering drugs, years of data, and family history of gastrointestinal cancer.

Table 5| Bisphosphonate use in gastric cancer cases and controls, numbers and odds ratios (95% confidence intervals) compared to non-use by database

	QResearch			CPRD			Combined analysis	
	Cases/controls	Adjusted* odds ratio (95% CI)	P value	Cases/controls	Adjusted* odds ratio (95% CI)	P value	Pooled odds ratio (95% CI)	P value
Total	3155/14 715	—	—	3157/14 686	—	—	—	—
Overall use								
None	3014/14 135	—	—	3018/13 992	—	—	—	—
At least 1 prescription	141/580	1.12 (0.87 to 1.44)	0.4	139/694	0.79 (0.62 to 1.01)	0.06	0.93 (0.78 to 1.11)	0.4
Regimen of use								
Daily	56/226	1.18 (0.84 to 1.67)	0.3	38/247	0.60 (0.41 to 0.87)	0.008	0.87 (0.67 to 1.12)	0.3
Weekly or monthly	85/354	1.07 (0.79 to 1.45)	0.6	101/447	0.91 (0.69 to 1.19)	0.5	0.98 (0.80 to 1.20)	0.8
Duration of use								
Short term (<1 year)	62/195	1.37 (0.98 to 1.90)	0.07	60/279	0.84 (0.61 to 1.15)	0.3	1.06 (0.84 to 1.33)	0.6
Long term (≥1 year)	79/385	0.96 (0.71 to 1.31)	0.8	79/415	0.75 (0.56 to 1.01)	0.06	0.85 (0.68 to 1.05)	0.1
1-6 months	33/109	1.32 (0.86 to 2.02)	0.2	37/170	0.85 (0.58 to 1.26)	0.4	1.04 (0.78 to 1.38)	0.8
7-36 months	73/285	1.13 (0.83 to 1.53)	0.4	76/332	0.90 (0.67 to 1.20)	0.5	1.00 (0.81 to 1.24)	1.0
37-72 months	23/131	0.83 (0.50 to 1.36)	0.5	20/138	0.53 (0.32 to 0.89)	0.02	0.67 (0.47 to 0.95)	0.03
>72 months	12/55	1.23 (0.62 to 2.41)	0.6	6/54	0.43 (0.18 to 1.03)	0.06	0.83 (0.49 to 1.41)	0.5
Trend for months of use	—	—	0.8	—	—	0.02	—	0.08
Alendronate								
Any use	102/361	1.47 (1.11 to 1.95)	0.008	100/457	0.93 (0.71 to 1.22)	0.6	1.16 (0.95 to 1.41)	0.1
<1 year	59/153	1.91 (1.34 to 2.72)	<0.001	47/212	0.94 (0.66 to 1.34)	0.7	1.34 (1.05 to 1.73)	0.02
≥1 year	43/208	1.08 (0.74 to 1.59)	0.7	53/245	0.93 (0.66 to 1.31)	0.7	0.99 (0.77 to 1.28)	1.0
Trend for months of use	—	—	0.6	—	—	0.1	—	0.5
Etidronate								
Any use	49/222	0.93 (0.65 to 1.32)	0.7	31/212	0.59 (0.39 to 0.88)	0.01	0.76 (0.58 to 0.99)	0.05
<1 year	26/87	1.20 (0.75 to 1.94)	0.4	15/97	0.62 (0.35 to 1.08)	0.09	0.91 (0.63 to 1.31)	0.6
≥1 year	23/135	0.76 (0.47 to 1.22)	0.3	16/115	0.56 (0.33 to 0.98)	0.04	0.67 (0.47 to 0.96)	0.03
Trend for months of use	—	—	0.6	—	—	0.04	—	0.1
Risedronate								
Any use	29/137	0.84 (0.54 to 1.30)	0.4	29/149	0.83 (0.54 to 1.27)	0.4	0.83 (0.61 to 1.13)	0.2
<1 year	16/67	0.89 (0.50 to 1.60)	0.7	13/74	0.75 (0.41 to 1.38)	0.4	0.82 (0.54 to 1.25)	0.4
≥1 year	13/70	0.72 (0.39 to 1.35)	0.3	16/75	0.91 (0.51 to 1.61)	0.7	0.82 (0.54 to 1.25)	0.3
Trend for months of use	—	—	0.5	—	—	0.7	—	0.5

*Adjusted for body mass index, smoking status, alcohol consumption, ethnicity, rheumatoid arthritis, osteoporosis and fractures, use of other osteoporosis drugs, vitamin D, NSAIDs, corticosteroids, acid lowering drugs, years of data, for family history of gastrointestinal cancer.

Table 6| Bisphosphonate use in cases of colorectal cancer and controls, numbers and odds ratios (95% confidence intervals) compared with non-use by database

	QResearch			CPRD			Combined analysis	
	Cases/controls	Adjusted* odds ratio (95% CI)	P value	Cases/controls	Adjusted* odds ratio (95% CI)	P value	Pooled odds ratio (95% CI)	P value
Total	20 106/93 954	—	—	19 035/89 111	—	—	—	—
Overall use								
None	19 177/89 609	—	—	18 133/84 880	—	—	—	—
At least 1 prescription	929/4345	1.03 (0.94 to 1.14)	0.5	902/4231	1.10 (1.00 to 1.22)	0.05	1.07 (1.00 to 1.14)	0.07
Regimen of use								
Daily	337/1607	1.01 (0.88 to 1.15)	0.9	294/1430	1.05 (0.91 to 1.21)	0.5	1.03 (0.93 to 1.13)	0.6
Weekly or monthly	592/2738	1.05 (0.93 to 1.18)	0.4	608/2801	1.13 (1.01 to 1.27)	0.03	1.09 (1.01 to 1.19)	0.03
Duration of use								
Short term (<1 year)	345/1599	1.02 (0.90 to 1.17)	0.7	358/1558	1.17 (1.03 to 1.33)	0.02	1.10 (1.00 to 1.20)	0.06
Long term (≥1 year)	584/2746	1.04 (0.92 to 1.17)	0.5	544/2673	1.06 (0.94 to 1.19)	0.4	1.05 (0.97 to 1.14)	0.3
1-6 months	212/993	1.02 (0.87 to 1.20)	0.8	252/1010	1.27 (1.09 to 1.48)	0.002	1.15 (1.03 to 1.28)	0.02
7-36 months	426/1982	1.03 (0.91 to 1.17)	0.6	394/1957	1.03 (0.91 to 1.17)	0.6	1.03 (0.95 to 1.13)	0.5
37-72 months	193/957	0.99 (0.83 to 1.18)	0.9	176/891	1.03 (0.86 to 1.23)	0.8	1.01 (0.89 to 1.14)	0.8
>72 months	98/413	1.19 (0.94 to 1.52)	0.1	80/373	1.13 (0.87 to 1.46)	0.3	1.16 (0.98 to 1.39)	0.09
Trend for months of use	—	—	0.3	—	—	0.5	—	0.2
Alendronate								
Any use	641/2894	1.10 (0.98 to 1.22)	0.1	627/2931	1.10 (0.98 to 1.22)	0.1	1.10 (1.01 to 1.19)	0.02
<1 year	305/1331	1.13 (0.98 to 1.30)	0.1	283/1264	1.12 (0.97 to 1.30)	0.1	1.13 (1.02 to 1.25)	0.02
≥1 year	336/1563	1.07 (0.93 to 1.23)	0.3	344/1667	1.07 (0.93 to 1.23)	0.3	1.07 (0.97 to 1.18)	0.2
Trend for months of use	—	—	0.2	—	—	0.5	—	0.2
Etidronate								
Any use	303/1575	0.90 (0.78 to 1.03)	0.1	271/1332	1.00 (0.86 to 1.15)	1.0	0.94 (0.86 to 1.04)	0.3
<1 year	114/617	0.85 (0.69 to 1.05)	0.1	119/569	1.02 (0.83 to 1.26)	0.8	0.93 (0.81 to 1.08)	0.4
≥1 year	189/958	0.93 (0.79 to 1.10)	0.4	152/763	0.98 (0.82 to 1.18)	0.8	0.95 (0.84 to 1.08)	0.5
Trend for months of use	—	—	0.9	—	—	0.6	—	0.7
Risedronate								
Any use	194/928	1.00 (0.85 to 1.18)	1.0	198/880	1.13 (0.96 to 1.34)	0.2	1.06 (0.95 to 1.20)	0.3
<1 year	80/442	0.87 (0.68 to 1.12)	0.3	104/412	1.27 (1.02 to 1.59)	0.04	1.07 (0.91 to 1.27)	0.4
≥1 year	114/486	1.11 (0.90 to 1.38)	0.3	94/468	1.00 (0.79 to 1.26)	1.0	1.06 (0.90 to 1.24)	0.5
Trend for months of use	—	—	0.7	—	—	0.8	—	0.9

*Adjusted for, BMI, smoking status, alcohol consumption, ethnicity, rheumatoid arthritis, osteoporosis and fractures, use of other osteoporosis drugs, vitamin D, NSAIDs, corticosteroids, acid lowering drugs, years of data, and for family history of colorectal cancer, diabetes, colitis, and Crohn's disease.

Table 7| **Sensitivity analyses by definition of length of use of bisphosphonates (short term (≤1 year) and long term (>1 year)) and varying definitions of use**

	QResearch			CPRD		
Term of use	Exposed cases	Adjusted* odds ratio (95% CI)	P value	Exposed cases	Adjusted* odds ratio (95% CI)	P value
Oesophageal cancer						
Use defined as at least 2 prescriptions in observation period						
≤1 year	74	1.03 (0.77 to 1.36)	0.9	65	0.88 (0.65 to 1.18)	0.4
>1 year	158	0.92 (0.73 to 1.16)	0.5	173	1.23 (0.98 to 1.54)	0.07
Use defined as at least 1 prescription before index date						
≤1 year	111	1.23 (0.96 to 1.56)	0.1	103	1.17 (0.91 to 1.50)	0.2
>1 year	171	0.94 (0.75 to 1.18)	0.6	182	1.23 (0.99 to 1.53)	0.06
Use defined as at least 1 prescription between 72 and 6 months before index date in patients with at least 6 years of records						
≤1 year	75	0.99 (0.74 to 1.32)	0.9	78	1.09 (0.81 to 1.45)	0.6
>1 year	142	0.93 (0.73 to 1.18)	0.5	145	1.26 (0.98 to 1.60)	0.07
Gastric cancer						
Use defined as at least 2 prescriptions in observation period						
≤1 year	53	1.49 (1.04 to 2.13)	0.03	50	0.75 (0.53 to 1.06)	0.1
>1 year	79	0.97 (0.71 to 1.31)	0.8	79	0.73 (0.55 to 0.99)	0.04
Use defined as at least 1 prescription before the index date						
≤1 year	61	1.17 (0.84 to 1.62)	0.3	70	1.02 (0.76 to 1.38)	0.9
>1 year	88	0.93 (0.69 to 1.25)	0.6	90	0.82 (0.61 to 1.08)	0.2
Use defined as at least 1 prescription between 72 and 6 months before index date in patients with at least 6 years of records						
≤1 year	54	1.21 (0.84 to 1.75)	0.3	52	0.85 (0.60 to 1.21)	0.4
>1 year	66	0.90 (0.64 to 1.26)	0.5	65	0.72 (0.52 to 0.99)	0.04
Colorectal cancer						
Use defined as at least 2 prescriptions in observation period						
≤1 year	273	1.02 (0.88 to 1.18)	0.8	291	1.09 (0.94 to 1.25)	0.3
>1 year	584	1.04 (0.92 to 1.16)	0.5	544	1.04 (0.92 to 1.16)	0.6
Use defined as at least 1 prescription before index date						
≤1 year	384	1.04 (0.92 to 1.18)	0.5	386	1.12 (0.99 to 1.27)	0.07
>1 year	633	1.02 (0.91 to 1.14)	0.7	589	1.06 (0.94 to 1.18)	0.4
Use defined as at least 1 prescription between 72 and 6 months before index date in patients with at least 6 years of records						
≤1 year	298	1.02 (0.88 to 1.18)	0.8	285	1.13 (0.97 to 1.30)	0.1
>1 year	492	1.00 (0.88 to 1.13)	1.0	476	1.08 (0.96 to 1.23)	0.2

*Adjusted for BMI, smoking status, alcohol consumption, ethnicity, rheumatoid arthritis, osteoporosis and fractures, use of other osteoporosis drugs, vitamin D, NSAIDs, corticosteroids, acid lowering drugs, and years of data and for family history of colorectal cancer, diabetes, colitis, and Crohn's disease.

Paper 6

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of non-gastrointestinal cancers: series of nested case-control studies using QResearch and CPRD data. *British Journal of Cancer*. 2013;109:795-806.

FULL PAPER

BJC

British Journal of Cancer (2013), 1–12 | doi: 10.1038/bjc.2013.383

Keywords: bisphosphonates; case–control studies; osteoporosis/drug therapy; epidemiology

Exposure to bisphosphonates and risk of common non-gastrointestinal cancers: series of nested case–control studies using two primary-care databases

Y Vinogradova*[,1], C Coupland[1] and J Hippisley-Cox[1]

[1]*Division of Primary Care, 13th floor, Tower Building, University Park, Nottingham NG2 7RD, UK*

Background: Bisphosphonates are the most commonly prescribed osteoporosis drugs but long-term effects are unclear, although antitumour properties are known from preclinical studies.

Methods: Nested case–control studies were conducted to investigate bisphosphonate use and risks of common non-gastrointestinal cancers (breast, prostate, lung, bladder, melanoma, ovarian, pancreas, uterus and cervical). Patients 50 years and older, diagnosed with primary cancers between 1997 and 2011, were matched to five controls using the UK practice-based QResearch and Clinical Practice Research Datalink (CPRD) databases. The databases were analysed separately and the results combined.

Results: A total of 91 556 and 88 845 cases were identified from QResearch and CPRD, respectively. Bisphosphonate use was associated with reduced risks of breast (odds ratio (OR): 0.92, 95% confidence interval (CI): 0.87–0.97), prostate (OR: 0.87, 95% CI: 0.79–0.96) and pancreatic (OR: 0.79, 95% CI: 0.68–0.93) cancers in the combined analyses, but no significant trends with duration. For alendronate, reduced risk associations were found for prostate cancer in the QResearch (OR: 0.81, 95% CI: 0.70–0.93) and combined (OR: 0.84, 95% CI: 0.75–0.93) analyses (trend with duration *P*-values 0.009 and 0.001). There were no significant associations from any of the other analyses.

Conclusion: In this series of large population-based case–control studies, bisphosphonate use was not associated with increased risks for any common non-gastrointestinal cancers.

Bisphosphonates were introduced as a treatment for postmenopausal osteoporosis (National Institute for Health and Clinical Excellence, 2008a, b) in the 1990s, and prescribing has substantially increased (Kanis *et al*, 2008). The effects of bisphosphonates are long term as the drugs accumulate in bones and are released for several years after treatment ends (Watts and Diab, 2010). It is a relatively new treatment and very few studies have looked at the long-term effect of bisphosphonates on risks of different cancers in the general population.

Although antitumour properties of bisphosphonates have been discovered in preclinical studies (Croucher *et al*, 2003; Guise, 2008) and reaffirmed in the treatment of bone metastases (Gnant, 2010), no long-term randomised clinical trials have been run to determine the effect of bisphosphonates on cancer incidence. Epidemiological studies have consistently reported a reduced risk of breast cancer in bisphosphonate users (Chlebowski *et al*, 2010; Newcomb *et al*, 2010; Rennert *et al*, 2010; Vestergaard *et al*, 2011; Cardwell *et al*, 2012), but the effects of bisphosphonates on other common non-gastrointestinal cancers are still uncertain, having been investigated in only one up-to-date observational study using the Clinical Practice Research Datalink (CPRD) (Cardwell *et al*, 2012). Previous studies have generally been insufficiently powered to

*Correspondence: Y Vinogradova; E-mail: yana.vinogradova@nottingham.ac.uk

Received 16 April 2013; revised 5 June 2013; accepted 11 June 2013

detect associations for other types of cancer, and few have looked at associations with individual types of bisphosphonate drugs, and so there is little data to establish definitive conclusions. Our aim, therefore, was to investigate the associations between bisphosphonates and risks of common cancers in the general population using a nested case–control design and including all available data from two large primary-care databases – both the QResearch and the CPRD.

MATERIALS AND METHODS

Data source. The two largest primary-care databases in the UK, QResearch and CPRD, were used. Each database covers around 6% of the UK population from more than 600 general practices, and contains electronic records including patient demographics, referrals, tests and prescriptions. Both are representative of the general population in the United Kingdom, have been rigorously validated using other sources of information (Jick *et al*, 2003; Hippisley-Cox *et al*, 2004) and have been used for a range of safety studies involving commonly prescribed medications (Green *et al*, 2010; Hippisley-Cox and Coupland, 2010; Parker *et al*, 2010; Walker *et al*, 2011). The protocol for this study was published in Vinogradova *et al* (2012) and, although originally based on QResearch, a replicate study has also been conducted using CPRD to examine any possible differences between the two and further increase in the statistical power by running analyses to derive combined results from both data sources.

Study design. Open cohorts of patients were identified in each database: patients were aged 50 years and older and registered with the practice at some time during the study period between January 1997 and July 2011. For this paper, we selected the most common solid, non-gastrointestinal cancers (breast, prostate, lung, bladder, pancreatic, ovarian and melanoma) as the outcomes and identified incident cases from the cohorts. Less common female cancers (cervix and uterus) were also considered. Each case was matched to up to five controls by age, sex, practice and calendar year. All controls were alive and registered with the practice at the date of the first recorded diagnosis of cancer in their matched case, which we defined as the index date for each case and their matched controls. For cases and controls, patients were included only if they had at least 2 years of data before their index date. Cases and controls with bisphosphonate prescriptions licensed for any malignancies before the index date (date of diagnosis for cases or equivalent date for controls) were excluded. For breast cancer, male patients and patients with a record of mastectomy before their first prescription of bisphosphonates were excluded. Patients with Paget's disease were also excluded.

Exposure to bisphosphonates. Exposure to bisphosphonates was assessed, including prescriptions for alendronate, etidronate, ibandronate and risedronate as the nationally licensed drugs for the treatment of osteoporosis (BNF 6.6.2) (British Medical Association and Royal Pharmaceutical Society of Great Britain, 2008). Information was extracted on all prescriptions for bisphosphonates during the observation period – defined as a period between the date of patient registration with the general practice and 6 months before the index date. Prescriptions in the past 6 months before the index date were not used to reduce protopathic bias because early symptoms of cancer such as low weight or bone ache could lead to bisphosphonate treatments. For the main analyses, a patient was considered to be a bisphosphonate user if they had at least one prescription during the observation period. Cumulative exposure was estimated by summing the durations of all bisphosphonate prescriptions for each patient, considering gaps of fewer than 90 days between two prescriptions as continuous therapy. Duration of exposure to bisphosphonates was analysed using the following categorisations: no use; short-term use (used for <1 year); long-term use (used for >1 year). For analyses of the most common cancers (breast, prostate, lung, bladder and melanoma), finer categories for duration of use of bisphosphonates were considered: no use, <6 months; 7–36 months; 37–72 months; and 73 months or more. A test for trend was performed using the actual number of months.

Confounding variables. All the analyses included potential confounders established as risk factors for cancer. Body mass index (BMI) (Henderson and Bernstein, 2008), a continuous variable, was based on values recorded at the date closest to 1 year before the index date. Using Read codes, smoking status (Hecht, 2008) (current smoker – light (1–9 cigarettes per day), medium (10–19), heavy (20 or more), ex-smoker, non-smoker); alcohol consumption (Schütze *et al*, 2011); and ethnicity (Ferlay *et al*, 2010) (White or not recorded, Black, Asian, Other) used values recorded at the closest date before the index date. The analysis also adjusted for history of osteoporosis (McGlynn *et al*, 2008), including diagnosis of osteoporosis or osteopenia or previous fractures (recorded before the index date); use of drugs increasing risks of fracture and cancer such as systemic corticosteroids and acid-suppressive medications (including H2 antagonists (BNF 1.3.1), proton pump inhibitors (BNF 1.3.5) and antacids (BNF 1.1.1)); (Corley *et al*, 2010) use of anti-inflammatory drugs (Coussens and Werb, 2002) (traditional non-steroidal anti-inflammatory drugs, cyclo-oxygenase 2 inhibitors and aspirin); (Gonzalez-Perez *et al*, 2003) and use of vitamin D (Mocellin, 2011) if they were prescribed at least 1 year before the index date. For female cancer patients, use of hormone replacement therapy and oral contraceptive pills were also added to the analyses.

If they were diagnosed at least a year before the index date, comorbidities, which affect risks of cancer, were also included: rheumatoid arthritis (Thomas *et al*, 2000) for any cancer; benign breast disease for breast cancer; diabetes for pancreatic (Vincent *et al*, 2011) and uterine (Burbos *et al*, 2010) cancers; hypertension for uterine cancer (Bangalore *et al*, 2011); and gastrointestinal disorders for pancreatic cancer (Vinogradova *et al*, 2012). The results were also adjusted for cancer-specific family histories of cancer (Mai *et al*, 2011) (to reduce recall bias in cases only if recorded at least 6 months before the index date; Chang *et al*, 2006).

Statistical analysis. This study used conditional logistic regression to estimate odds ratios with 95% CIs for cancers at each selected site. The Wald test was used for estimating the effects of duration and testing for differences between bisphosphonate types. Missing values for the confounding factors (BMI, smoking status and alcohol intake) were imputed using ICE multiple imputation program in Stata (StataCorp LP, College Station, TX, USA) (Royston, 2005) where all the confounding variables and exposure to bisphosphonates were included into the models. Ten imputed data sets were created and the results were combined using Rubin's rules (Royston, 2004).

Each database was analysed separately and the results combined using the method of Mantel and Haenszel for fixed-effect models.

The primary analyses were based on bisphosphonate exposure excluding prescriptions in the 6 months before the index date. Five sensitivity analyses were carried out for each data set. The first was to eliminate possible bias by redefining bisphosphonate exposure as at least two prescriptions and considering patients with one prescription only as non-users. Such patients might never have started treatment or have soon stopped it because of adverse gastrointestinal effects. The second sensitivity analysis was based on all prescriptions before the index date including the ones issued in the past 6 months. This aimed to eliminate another form of bias caused by a possible oversampling of unexposed cases and controls in the main analyses. The third sensitivity analysis was run on

patients with at least 6 years of medical records and redefined exposure based on prescription information only between 72 and 6 months before the index date. This was to eliminate possible bias arising from the different observation times for patients in the main analyses.

Townsend scores, a measure of deprivation, were available for only 49% of CPRD practices, and only 36% of CPRD patients had it recorded. Deprivation, therefore, was not included as a confounding variable in any of the main analyses. The fourth sensitivity analysis included Townsend scores as a confounding variable, but was run only on patients with a valid code. The fifth and final sensitivity analysis was run on observations with recorded values for BMI, smoking status and alcohol consumption. For the fourth and fifth sensitivity analyses, the definitions for use of bisphosphonates and years of medical records were identical to the main analysis.

Although sample size calculations were carried out and presented in the protocol (Vinogradova *et al*, 2012), all available data were used in the analyses. We considered a 1% level as statistically significant to allow for multiple comparisons, but have presented 95% CIs in our results to create parity with other studies. Stata Version 12 was used for all analyses.

RESULTS

Study population. Within the study period, we identified 91 556 and 88 845 cases of cancers of interest in people aged 50 years and older from QResearch and CPRD, respectively. These were matched to 427 674 and 415 583 controls, respectively, all with at least 2 years of medical records (Figure 1). Median years of available records was 17 (interquartile range 10–28 in QResearch cases and controls, 10–29 in CPRD cases and 10–30 CPRD controls), and it was consistent both for cancer sites and for cases and controls. Table 1 shows the numbers of cases and controls for each cancer site and also the characteristics of the study population. QResearch and CPRD provided similar samples from the general population. The QResearch cases and controls were about the same age (mean 69.4 years, standard deviation 9.8 in cases; 69.4 years, 9.7 in controls) as CPRD (69.9 years, 10.4 in cases; 69.7 years, 10.2 in controls) but included fewer women (48% *vs* 51%), and so had slightly different incidences in cancers of interest.

The proportion of bisphosphonate users (4.2%) was the same in both databases and in cases and controls, with similar proportions of patients prescribed different types of bisphosphonates. Figure 2 shows that in both databases the proportion of cases and controls with at least one bisphosphonate prescription consistently grew from 1% in 1997 to 7% in 2011. Use of etidronate decreased but use of alendronate and risedronate increased, for alendronate reaching over 80% for bisphosphonate users with at least one prescription. Most bisphosphonate users were older than 60 years (92% in both databases) and their median duration of use was 20 months (interquartile range 7–43) in cases and controls in QResearch and 19 months (6–40 in cases and 6–41 in controls) in CPRD. The median duration varied slightly between the cancer sites, from 16.5 months for cervix cases and 27 for melanoma cases in QResearch, and 17 for lung cancer cases to 22.5 for melanoma cases in CPRD. A much higher proportion of women were users than men (6.6% women *vs* 2.0% men in cases, 6.8% women *vs* 1.7% men in controls for QResearch, 6.5% women *vs* 1.9% men in cases and 6.6% women *vs* 1.7% men in controls for CPRD). Use of bisphosphonates did not necessarily follow a recorded diagnosis of osteoporosis/osteopenia or a history of fractures; over 30% of bisphosphonate users did not have either factor recorded (38% cases and 36% controls in QResearch and 36% cases and 34% controls for CPRD).

Table 2 shows the number and proportion of cases and controls who were bisphosphonate users for each cancer site and overall use of the drug by database. Figure 3 contains the results of combined analyses for overall bisphosphonate use. Table 3 presents trend test results along with odds ratios for short- and long-term use for the seven most common cancers. Table 4 shows further analyses for the associations of different types of bisphosphonate use with breast, prostate and lung cancer risks.

Breast cancer. Overall use of bisphosphonates was associated with a reduced risk of breast cancer but it was only significant in QResearch analyses (adjusted odds ratio (AOR): 0.89, 95% CI: 0.82–0.97) and the combined analyses (AOR: 0.92, 95% CI: 0.87–0.98). Clinical Practice Research Datalink analyses demonstrated a similar direction of the association (AOR: 0.95, 95% CI: 0.88–1.03), but it was not statistically significant ($P = 0.2$). None of the associations were duration-dependent, with the only statistically significantly decreased risk for QResearch ($P = 0.008$) in the subcategory between 7 and 36 months (AOR: 0.87, 95% CI: 0.78–0.96 for QResearch; 0.95, 0.86–1.05 for CPRD; 0.91, 0.85–0.98 for combined). The risks did not vary between bisphosphonate types and none of them had statistically significant associations with breast cancer risk.

The sensitivity analysis, which defined bisphosphonate use as at least two prescriptions, demonstrated an even stronger reduced association with breast cancer risk (AOR: 0.87, 95% CI: 0.80–0.94, $P < 0.001$ for QResearch; 0.94, 0.87–1.01, $P = 0.1$ for CPRD; 0.90, 0.85–0.96, $P < 0.001$ for combined), but it remained statistically significant only for short-term use, and only for QResearch and combined (AOR: 0.80, 95% CI: 0.71–0.91, $P < 0.001$ for QResearch; 0.87, 0.80–0.94, $P < 0.001$ for combined) but not for CPRD (AOR: 0.92, 95% CI: 0.82–.03). There was no significant association with long-term use for any analyses.

Other female cancers. Although QResearch analyses for ovarian cancer showed an almost 20% increased, but not statistically significant, risk (AOR: 1.19, 95% CI: 0.96–1.47), CPRD analyses had an opposite, and also not statistically significant, association (AOR: 0.84, 95% CI: 0.67–1.04), with combined results showing no association at all (AOR: 1.00, 95% CI: 0.86–1.16). None of the results from the different databases were duration-dependent.

There was a significant association between risedronate use and ovarian cancer risk for QResearch (24 exposed cases) and the combined analyses (56 exposed cases) (AOR: 0.48, 95% CI: 0.31–0.75, $P = 0.001$ for QResearch; 0.62, 0.46–0.84, $P = 0.002$ for combined). The CPRD analysis (32 exposed cases) showed a reduced risk but not statistically significantly (AOR: 0.77, 95% CI: 0.52–1.16). The trend tests were not statistically significant ($P = 0.013$ for QResearch; $P = 0.6$ for CPRD; $P = 0.07$ for combined) and this was consistent across all sensitivity analyses.

Overall use of bisphosphonates was not associated with cervical or uterine cancer risk for either database or for the combined analyses. Further analyses demonstrated no associations with duration of use.

Prostate cancer. Adjusted analyses in both databases demonstrated a reduced risk associated with bisphosphonate use but not significant (AOR: 0.90, 95% CI: 0.79–1.02 for QResearch; 0.84, 0.73–0.96 for CPRD), but contributing to a significant association in the combined analysis (AOR: 0.87; 95% CI: 0.79–0.96). Although neither of the databases showed a significant duration-dependent association (QResearch $P_{\text{trend}} = 0.064$; CPRD $P_{\text{trend}} = 0.03$), the combined analyses did ($P_{\text{trend}} = 0.005$), with a 15% reduced risk associated with long-term use of bisphosphonates (AOR: 0.85,; 95% CI: 0.76–0.95).

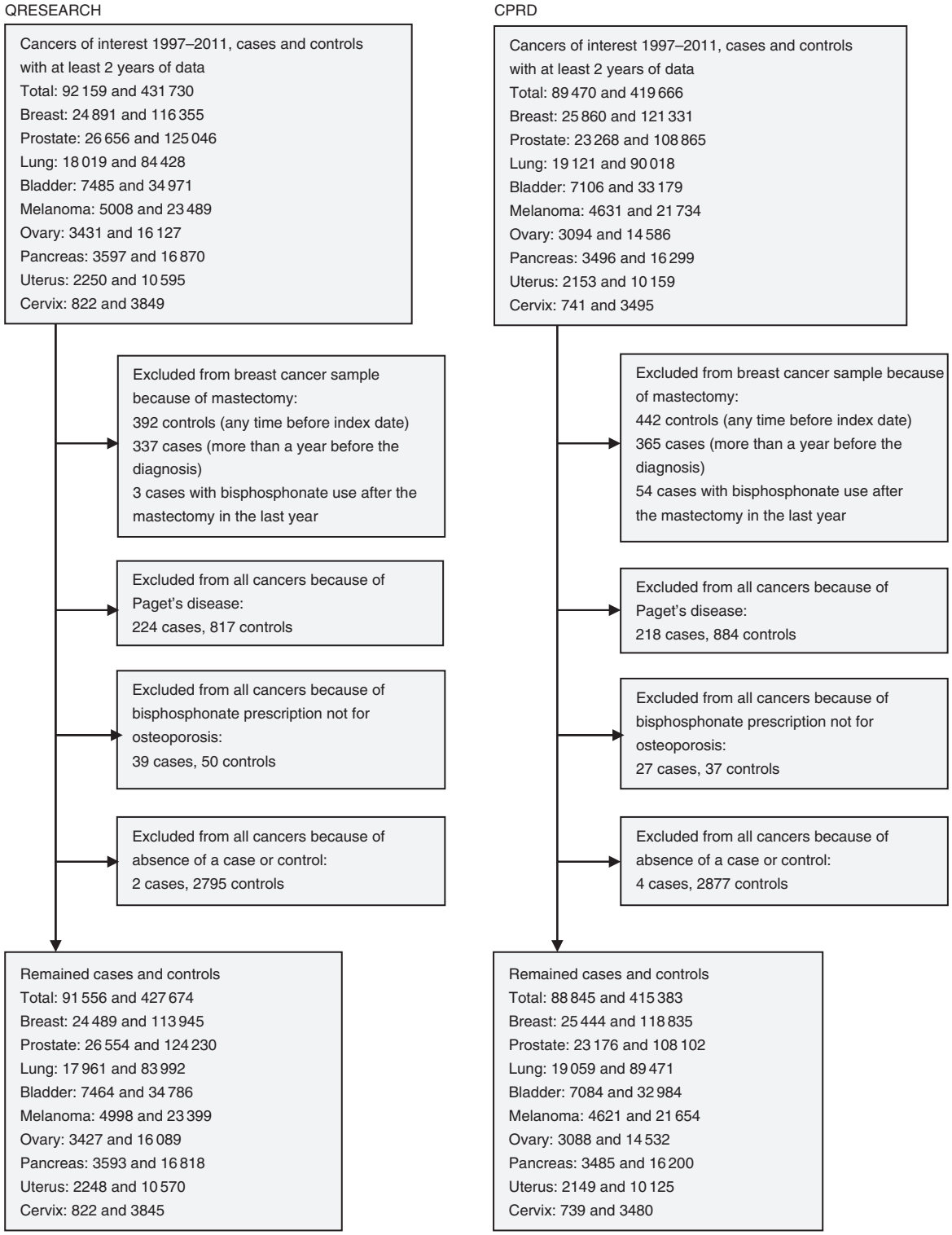

Figure 1. **Flow of the included patients for QResearch and CPRD analyses.**

Further analyses showed significantly reduced risks for alendronate users in QResearch, with a greater reduced risk for long-term users (AOR: 0.75; 95% CI: 0.62–0.92), which was duration-dependent ($P_{trend} = 0.009$). Clinical Practice Research Datalink associations were directionally similar to QResearch but not significant and also with lower risk for long-term users (AOR: 0.86; 95% CI: 0.70–1.05) and not duration-dependent ($P_{trend} = 0.05$), although contributing to a significant association (AOR: 0.80; 95% CI: 0.70–0.92) and duration dependency ($P_{trend} = 0.001$) in the combined analysis.

All sensitivity analyses with other definitions of use were consistent in showing a reduced risk in long-term alendronate users apart from the one including prescriptions from the past 6 months before the index date. This analysis also demon-

Table 1. Baseline characteristics in cases and all matched controls by database (QResearch or CPRD)

	QResearch		CPRD	
	Cases	**Controls**	**Cases**	**Controls**
Total	91 556	427 674	88 845	415 383
Breast	24 489	113 945	25 444	118 835
Prostate	26 554	124 230	23 176	108 102
Lung	17 961	83 992	19 059	89 471
Bladder	7464	34 786	7084	32 984
Skin	4998	23 399	4621	21 654
Ovary	3427	16 089	3088	14 532
Pancreas	3593	16 818	3485	16 200
Uterus	2248	10 570	2149	10 125
Cervix	822	3845	739	3480
Sex				
Male	47 193 (51.5)	220 492 (51.6)	43 596 (49.1)	203 487 (49.0)
Female	44 363 (48.5)	207 182 (48.4)	45 249 (50.9)	211 896 (51.0)
Age band (years)				
30–54	7345 (8.0)	34 000 (7.9)	6998 (7.9)	32 749 (7.9)
55–64	22 786 (24.9)	106 266 (24.8)	22 071 (24.8)	104 095 (25.1)
65–74	30 410 (33.2)	143 528 (33.6)	28 407 (32.0)	134 902 (32.5)
75–84	26 177 (28.6)	122 798 (28.7)	23 864 (26.9)	112 533 (27.1)
85 +	4838 (5.3)	21 082 (4.9)	7505 (8.4)	31 104 (7.5)
Ethnicity				
White	21 477 (23.5)	90 874 (21.2)	7034 (7.9)	30 294 (7.3)
Not recorded[a]	69 131 (75.5)	331 241 (77.5)	78 841 (88.7)	371 491 (89.4)
Non-white	948 (1.0)	5559 (1.3)	2970 (3.3)	13 598 (3.3)
Asian[b]	364 (0.4)	2726 (0.6)	194 (0.2)	1516 (0.4)
Black[b]	397 (0.4)	1834 (0.4)	166 (0.2)	649 (0.2)
Other[b]	187 (0.2)	999 (0.2)	2610 (2.9)	11 433 (2.8)
Deprivation, Townsend quintile				
1, most affluent[c]	23 442 (26.3)	110 919 (26.7)	9753 (30.5)	45 736 (30.6)
2[c]	19 959 (22.4)	94 167 (22.7)	7695 (24.0)	36 203 (24.2)
3[c]	18 032 (20.2)	84 427 (20.3)	6253 (19.5)	29 740 (19.9)
4[c]	15 768 (17.7)	72 202 (17.4)	5007 (15.6)	23 075 (15.4)
5, most deprived[c]	12 075 (13.5)	53 799 (12.9)	3290 (10.3)	14 914 (10.0)
Not recorded	2280 (2.5)	12 160 (2.8)	56 847 (64.0)	265 715 (64.0)
BMI (kg m^{-2})				
15–24[c]	27 225 (38.0)	120 028 (37.3)	29 310 (39.1)	131 726 (38.3)
25–29[c]	28 956 (40.4)	130 874 (40.7)	29 808 (39.7)	138 713 (40.3)
30–49[c]	15 409 (21.5)	70 904 (22.0)	15 913 (21.2)	73 554 (21.4)
Not recorded	19 966 (21.8)	105 868 (24.8)	13 814 (15.5)	71 390 (17.2)
Smoking status				
Non-smoker[c]	34 949 (41.4)	182 349 (48.0)	40 128 (47.6)	21 2180 (55.4)
Ex-smoker[c]	31 475 (37.3)	136 575 (36.0)	26 958 (32.0)	112 142 (29.3)
Current light smoker[c]	5633 (6.7)	20 787 (5.5)	3966 (4.7)	13 871 (3.6)
Current moderate smoker[c]	7825 (9.3)	26 747 (7.0)	8838 (10.5)	30 846 (8.1)
Current heavy smoker[c]	4467 (5.3)	13 187 (3.5)	4382 (5.2)	13 646 (3.6)
Not recorded	7207 (7.9)	48 029 (11.2)	4573 (5.1)	32 698 (7.9)
Use of alcohol				
No use[c]	22 061 (29.9)	101 997 (30.5)	19 714 (25.8)	92 647 (26.6)
Ex use[c]			2030 (2.7)	7672 (2.2)
Light[c]	44 809 (60.7)	203 342 (60.8)	44 448 (58.1)	204 556 (58.8)
Moderate and more[c]	7009 (9.5)	29 036 (8.7)	10 250 (13.4)	42 911 (12.3)
Not recorded	17 677 (19.3)	93 299 (21.8)	12 403 (14.0)	67 597 (16.3)

Table 1. (Continued)

	QResearch		CPRD	
	Cases	Controls	Cases	Controls
Comorbidities				
Upper GI	19 250 (21.0)	83 377 (19.5)	24 302 (27.4)	107 660 (25.9)
Diabetes	8140 (8.9)	36 769 (8.6)	7990 (9.0)	37 011 (8.9)
Pancreatitis	492 (0.5)	1981 (0.5)	552 (0.6)	2102 (0.5)
Hypertension	31 466 (34.4)	142 577 (33.3)	31 169 (35.1)	142 474 (34.3)
Chronic kidney disease	3444 (3.8)	15 558 (3.6)	3710 (4.2)	16 425 (4.0)
Rheumatoid arthritis	1424 (1.6)	6591 (1.5)	1673 (1.9)	7310 (1.8)
Benign breast disease	2727 (3.0)	10 077 (2.4)	2761 (3.1)	9653 (2.3)
History of osteoporosis				
Osteoporosis/osteopenia	3682 (4.0)	16 998 (4.0)	4023 (4.5)	18 474 (4.4)
Osteoporotic fractures	3322 (3.6)	16 013 (3.7)	2375 (2.7)	10 342 (2.5)
Family history of cancer	2841 (3.1)	11 756 (2.7)	2836 (3.2)	11 364 (2.7)
Medications (excluding past 6 months)				
Acid-lowering drugs	32 523 (35.5)	140 637 (32.9)	33 531 (37.7)	145 409 (35.0)
NSAIDs	52 422 (57.3)	233 143 (54.5)	55 106 (62.0)	248 518 (59.8)
Corticosteroids	14 156 (15.5)	57 106 (13.4)	13 908 (15.7)	54 233 (13.1)
Calcium	1023 (1.1)	4816 (1.1)	6286 (7.1)	28 984 (7.0)
Vitamin D	4891 (5.3)	22 582 (5.3)	4960 (5.6)	23 035 (5.5)
Hormone replacement[d]	12 071 (13.2)	51 861 (12.1)	12 633 (14.2)	55 666 (13.4)
Oral contraceptives[d]	773 (0.8)	3551 (0.8)	1236 (1.4)	5490 (1.3)
Bisphosphonates				
Any	3827 (4.2)	17 883 (4.2)	3769 (4.2)	17 490 (4.2)
Alendronate	2589 (2.8)	12 010 (2.8)	2538 (2.9)	11 969 (2.9)
Etidronate	1286 (1.4)	5949 (1.4)	1170 (1.3)	5471 (1.3)
Risedronate	851 (0.9)	3985 (0.9)	806 (0.9)	3715 (0.9)
Ibandronate	74 (0.1)	435 (0.1)	75 (0.1)	353 (0.1)
Other osteoporosis drugs				
Any	319 (0.3)	1641 (0.4)	333 (0.4)	1685 (0.4)
Raloxifen	192 (0.2)	1032 (0.2)	217 (0.2)	1081 (0.3)
Strontium	95 (0.1)	456 (0.1)	97 (0.1)	509 (0.1)
Calcitonin	38 (0.0)	207 (0.0)	34 (0.0)	162 (0.0)

Abbreviations: CPRD = Clinical Practice Research Datalink; GI = gastrointestinal; NSAID = non-steroidal anti-inflammatory drugs. Values are numbers and %.
[a]Assumed White for the analyses.
[b]Broken down categories for non-White group.
[c]Proportion only within recorded data.
[d]Only for women cancers.

strated an increased risk for very short-term use (up to 6 months) but only in QResearch (AOR: 1.29; 95% CI: 1.08–1.53, $P = 0.004$). As for overall bisphosphonate use, the sensitivity analysis for at least two prescriptions reached a statistically significant level in CPRD (AOR: 0.83, 95% CI: 0.72–0.95, $P = 0.008$).

Lung cancer. After adjusting for confounders, there was no significant association with bisphosphonate use (AOR: 0.97, 95% CI: 0.88–1.08 in QResearch; 1.12, 1.01–1.23 in CPRD; 1.04, 0.97–1.12 for combined). Further analyses demonstrated no associations with duration or type of bisphosphonate.

Results were consistent in sensitivity analyses apart from one – the sensitivity analysis, which included prescriptions from the past 6 months before the index date showed increased risk associated with short-term use in both databases (AOR: 1.34, 95% CI: 1.18–1.52, $P < 0.001$ for QResearch; 1.30, 1.15–1.47, $P < 0.001$ for CPRD).

Bladder cancer. Adjusting for confounders showed no association between bisphosphonate use and risk of bladder cancer (AOR: 0.96, 95% CI: 0.80–1.14 for QResearch; 0.94, 0.79–1.12 for CPRD; 0.95, 0.84–1.08 for combined). Sensitivity analyses also showed no associations.

Melanoma. There was no association between overall bisphosphonate use and risk of melanoma in either database (AOR: 1.05, 95% CI: 0.87–1.28 for QResearch; 0.95, 0.77–1.19 for CPRD; 1.01, 0.87–1.17 for combined), and analyses did not show statistically significantly increased risk with longer use in either database or the combined ($P_{trend} = 0.013$ for QResearch; $P_{trend} = 0.5$ for CPRD; $P_{trend} = 0.02$ for combined). The results were consistent across the sensitivity analyses, except for the one defining use as at least one prescription between 6 and 72 months.

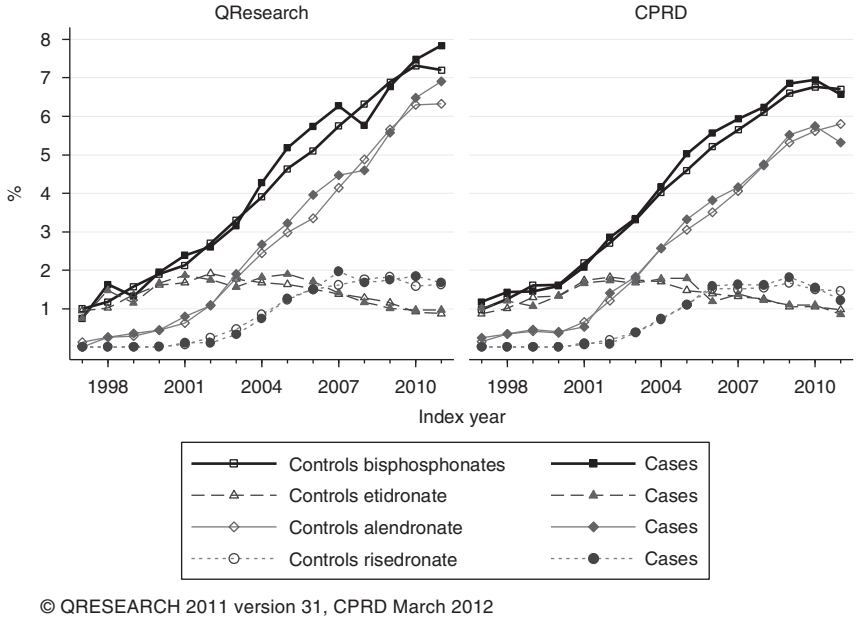

Figure 2. **Proportion of cases and controls in QResearch and CPRD with at least one prescription of bisphosphonate by index year.**

Table 2. Bisphosphonate use in cancer cases and controls, numbers and odds ratios (95% CIs) compared with non-use by database

	QResearch			CPRD			Combined analysis	
Cancer site	Cases/ controls	Adjusted* odds ratio (95% CI)	*P*-value	Cases/ controls	Adjusted* odds ratio (95% CI)	*P*-value	Pooled odds ratio (95% CI)	*P*-value
Breast[a]	1304/6923	0.89 (0.82–0.97)	0.005	1324/6847	0.95 (0.88–1.03)	0.2	0.92 (0.87–0.97)	0.004
Prostate[b]	460/2150	0.90 (0.79–1.02)	0.1	376/1901	0.84 (0.73–0.96)	0.012	0.87 (0.79–0.96)	0.003
Lung[c]	1035/3809	0.97 (0.88–1.08)	0.6	1114/3911	1.12 (1.01–1.23)	0.03	1.04 (0.97–1.12)	0.2
Bladder[d]	274/1219	0.96 (0.80–1.14)	0.6	280/1263	0.94 (0.79–1.12)	0.5	0.95 (0.84–1.08)	0.4
Melanoma[e]	241/1067	1.05 (0.87–1.28)	0.6	178/881	0.95 (0.77–1.19)	0.7	1.01 (0.87–1.17)	0.9
Ovary[f]	204/939	1.19 (0.96–1.47)	0.1	170/937	0.84 (0.67–1.04)	0.1	1.00 (0.86–1.16)	1.0
Pancreas[g]	178/886	0.81 (0.64–1.01)	0.06	196/918	0.78 (0.63–0.97)	0.03	0.79 (0.68–0.93)	0.003
Uterus[h]	99/671	1.07 (0.79–1.44)	0.7	96/636	0.95 (0.71–1.27)	0.7	1.00 (0.81–1.24)	1.0
Cervix[i]	32/219	0.78 (0.48–1.27)	0.3	35/196	1.21 (0.76–1.93)	0.4	0.98 (0.70–1.37)	0.9

Abbreviations: BMI = body mass index; CI = confidence interval; CPRD = Clinical Practice Research Datalink; GI = gastrointestinal; NSAID = nonsteroidal anti-inflammatory drugs. *Adjusted for BMI, smoking status, alcohol consumption, ethnicity, rheumatoid arthritis, osteoporosis and fractures, use of other osteoporosis drugs, vitamin D, NSAIDs, corticosteroids, acid-lowering drugs and years of data.

[a]Also adjusted for family history of breast cancer, benign breast disease, use of oral contraceptives and hormone replacement therapy.

[b]Also adjusted for family history of prostate cancer.

[c]Also adjusted for family history of lung cancer.

[d]Also adjusted for family history of cancer and chronic kidney disease.

[e]Also adjusted for family history of cancer.

[f]Also adjusted for family history of ovarian cancer, use of oral contraceptives and hormone replacement therapy.

[g]Also adjusted for family history of GI cancer, GI disease, diabetes and history of pancreatitis.

[h]Also adjusted for family history of cancer, hypertension, diabetes, use of oral contraceptives and hormone replacement therapy.

[i]Also adjusted for family history of cancer, use of oral contraceptives and hormone replacement therapy.

Here use of etidronate in the QResearch analysis was associated with increased risk of melanoma (AOR: 1.67, 95% CI: 1.23–2.26, $P<0.001$); however, this association was not duration-dependent ($P_{trend}=0.02$).

Pancreatic cancer. After adjusting for the confounders analyses in both databases showed a reduced risk of pancreatic cancer in bisphosphonate users, although it was not statistically significant (AOR: 0.81, 95% CI: 0.64–1.01 for QResearch; 0.78, 0.63–0.97 for CPRD). The combined analysis, however, demonstrated a statistically significant association (AOR: 0.79, 95% CI: 0.68–0.93). Short-term use was associated with an even lower but not statistically significant risk in all analyses (AOR: 0.78, 95% CI: 0.57–1.06 for QResearch; 0.77, 0.57–1.02 for CPRD; 0.77,

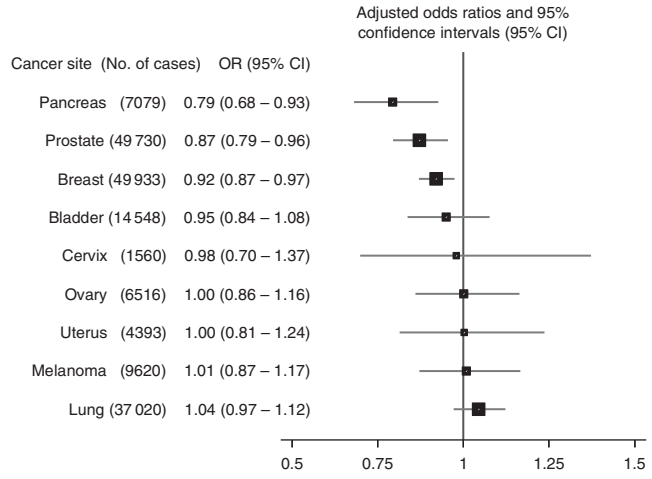

Figure 3. **Risk of cancer in patients prescribed bisphosphonates, combined results for QResearch and CPRD analyses.**

0.63–0.96 for combined) and there was also no trend for duration of use.

A sensitivity analysis defining use as at least two prescriptions reached statistically significant level for the CPRD analysis in overall use (0.73, 0.59–0.91, $P = 0.005$) and for short-term use (AOR: 0.65, 95% CI: 0.47–0.89, $P = 0.008$). The other sensitivity analyses did not demonstrate any significant associations.

Additional information. Sensitivity analyses on patients with valid BMI, smoking status and alcohol consumption data and on patients with valid Townsend codes showed similar results, which are available from the authors.

DISCUSSION

Summary. This study, based on medical records from the two largest UK primary-care databases, showed that the use of bisphosphonates was not associated with increased risks for any of the most common solid non-gastrointestinal cancers. Decreased risks of breast, prostate and pancreatic cancers had no duration relationship with bisphosphonate use and were found only in short-term users. Alendronate use, however, was associated with a decreased risk of prostate cancer and this association was duration-dependent.

Strengths and limitations. Based on the two largest primary-care databases, including more recent data than previous studies and covering the period when bisphosphonates have become much more widely prescribed in the general population, our study provides more data for investigating the duration effects of bisphosphonate use and over longer periods than previous studies. All eligible cancer cases and controls, alive or dead, were included into the study. This representativeness of the databases, and the lack of recall, selection and respondent biases, also makes the results more generalisable.

The limitations include possible uncertainties in cancer diagnosis. The selection of cases was based on the first record of a cancer and the exact origin site may have been determined only later, but this level of detail was not available across all records. A systematic review based on GPRD validation studies reported that, on average, 95% of diagnoses of cancer recorded on the GP electronic record were confirmed by other data sources (Herrett *et al*, 2010). Information about cancer stage and the results of histological investigations were also not consistently recorded across general practices and so were not used. Another limitation is the potential misclassification of bisphosphonate use. The analyses were based on prescriptions not on actual use and no data were available on adherence to medications. There is no reason, however, to suppose that non-adherence differs systematically between cases and controls.

Other limitations include information bias and missing data. Lack of symptom or family history records might arise simply because the patients have not reported them. Information on certain risk factors such as the level of physical activity, diet and cancer screening tests (mammography, prostate-specific antigen test) was also not reliably recorded, so these factors were not included in the analyses. Further, results of any bone mineral density tests were not consistently recorded and were not used in the analysis, and so there might be some residual confounding. Important confounders such as smoking or BMI had some missing data, so these had to be imputed.

Bisphosphonate users. Comparisons between bisphosphonate users and non-users reflect the recommendations for targeting the group with osteoporosis, and so some characteristics and comorbidities such as low BMI and rheumatoid arthritis in users were expected as these are risk factors for primary osteoporosis. As secondary osteoporosis is more likely to develop in patients with impaired digestion, use of acid-lowering drugs and also use of corticosteroids (Mauck and Clarke, 2006; Hippisley-Cox and Coupland, 2009) were much higher for bisphosphonate users.

Breast cancer. Reduced breast cancer risk associated with bisphosphonates use has been shown in five studies to date, all smaller than the current one (Chlebowski *et al*, 2010; Newcomb *et al*, 2010; Rennert *et al*, 2010; Vestergaard *et al*, 2011; Cardwell *et al*, 2012). No relation between risk and duration of use could suggest that osteoporosis is responsible for such reductions (Chen *et al*, 2008). Adjustments for osteoporosis, however, have demonstrated the independent effect of bisphosphonates. Reduced risk associated with recent, but not remote, use of bisphosphonates might also be explained by a possible bisphosphonate-related prevention of progress for undiagnosed cancers up to an invasive stage. This has already been shown by Chlebowski's study (Chlebowski *et al*, 2010) and supported by studies in menopausal women with early-stage breast cancer treated with bisphosphonates (Gnant, 2010). Although an Israeli (Rennert *et al*, 2010) case–control study showed a statistically significantly decreased risk associated with more than a year of bisphosphonate use (AOR: 0.61, 95% CI: 0.50–0.76), the study was subject to recall bias (participants reported their past use of bisphosphonates) and selection bias (all participants were alive at the moment of the interview). Another case–control study (Newcomb *et al*, 2010) with similar limitations in the design also reported a statistically significant trend ($P_{\text{trend}} = 0.01$) for duration of bisphosphonate use. A cohort study of Cardwell *et al* (2012) showed risk reduction associated with overall bisphosphonate use similar to ours (adjusted hazard ratio (AHR): 0.79, 95% CI: 0.62–1.01), and also becoming not statistically significant for patients with use of more than a year. Although our results for different types of bisphosphonate were in line with Vestergaard *et al*'s study (2011) for alendronate (AHR: 0.91, 95% CI: 0.81–1.03), the risk reduction for etidronate in our study did not reach a statistically significant level.

Prostate cancer. The association between overall use of bisphosphonates and prostate cancer risk was similar to that for breast cancer, also without statistically significant effects for therapy duration. Although the association between alendronate use and

Table 3. Bisphosphonate short- and long-term use in cancer cases and controls, numbers and odds ratios (95% CIs) compared with non-use by database

Cancer site	QResearch			CPRD			Combined analysis	
Terms of use	Cases/ controls	Adjusted* odds ratio (95% CI)	*P*-value	Cases/ controls	Adjusted* odds ratio (95% CI)	*P*-value	Pooled odds ratio (95% CI)	*P*-value
Breast	Trend		**0.5**	Trend		**0.3**	Trend	**0.2**
Short term (less than a year)	462/2523	0.86 (0.77–0.97)	0.01	508/2620	0.95 (0.86–1.06)	0.4	0.91 (0.84–0.98)	0.02
Long term (at least a year)	842/4400	0.91 (0.83–1.00)	0.05	816/4227	0.95 (0.86–1.04)	0.3	0.93 (0.87–0.99)	0.03
Prostate	Trend		**0.06**	Trend		**0.03**	Trend	**0.005**
Short term (less than a year)	186/838	0.95 (0.80–1.13)	0.6	149/764	0.84 (0.70–1.02)	0.07	0.90 (0.79–1.02)	0.1
Long term (at least a year)	274/1312	0.86 (0.74–1.00)	0.06	227/1137	0.84 (0.71–0.99)	0.04	0.85 (0.76–0.95)	0.005
Lung	Trend		**0.7**	Trend		**1.0**	Trend	**0.8**
Short term (less than a year)	415/1351	1.04 (0.90–1.19)	0.6	467/1507	1.15 (1.01–1.32)	0.03	1.10 (1.00–1.21)	0.06
Long term (at least a year)	620/2458	0.93 (0.82–1.05)	0.2	647/2404	1.09 (0.96–1.23)	0.2	1.01 (0.92–1.10)	0.9
Bladder	Trend		**0.3**	Trend		**0.6**	Trend	**0.3**
Short term (less than a year)	109/455	0.99 (0.79–1.26)	1.0	117/464	1.06 (0.84–1.33)	0.6	1.03 (0.87–1.21)	0.8
Long term (at least a year)	165/764	0.93 (0.76–1.15)	0.5	163/799	0.86 (0.69–1.06)	0.2	0.90 (0.77–1.04)	0.2
Melanoma	Trend		**0.01**	Trend		**0.5**	Trend	**0.02**
Short term (less than a year)	77/400	0.91 (0.70–1.20)	0.5	64/354	0.88 (0.65–1.18)	0.4	0.90 (0.73–1.10)	0.3
Long term (at least a year)	164/667	1.16 (0.92–1.45)	0.2	114/527	1.01 (0.78–1.31)	0.9	1.09 (0.92–1.29)	0.3
Ovaries	Trend		**0.4**	Trend		**0.1**	Trend	**0.7**
Short term (less than a year)	83/322	1.37 (1.04–1.81)	0.02	62/355	0.80 (0.59–1.08)	0.1	1.07 (0.87–1.31)	0.5
Long term (at least a year)	121/617	1.08 (0.84–1.38)	0.6	108/582	0.86 (0.66–1.11)	0.2	0.96 (0.81–1.15)	0.7
Pancreas	Trend		**0.2**	Trend		**0.05**	Trend	**0.02**
Short term (less than a year)	63/316	0.78 (0.57–1.06)	0.1	71/336	0.77 (0.57–1.02)	0.07	0.77 (0.63–0.96)	0.02
Long term (at least a year)	115/570	0.82 (0.63–1.06)	0.1	125/582	0.80 (0.62–1.02)	0.08	0.81 (0.67–0.97)	0.02

Abbreviations: CI = confidence interval; CPRD = Clinical Practice Research Datalink. *Adjusted for the confounders listed in the footnote for Table 2. Bold values indicate statistical significance.

decreased risk appeared to be significant even with a duration relationship, it was found only in the QResearch database and might again suggest a possible suppressing effect of alendronate on an already existing *in situ* cancer (Tuomela *et al*, 2008). In particular in the prostate, *in vitro* studies have observed substantial concentrations of nitrogen-containing bisphosphonates (zolendronic acid and ibandronate) after administration of the drug (Fournier *et al*, 2002), which could result in suppressed angiogenesis in prostate tumours. The only epidemiological study looking at prostate cancer with respect to bisphosphonate use has also demonstrated lower risk for bisphosphonate users, although not statistically significant (AHR: 0.71, 95% CI: 0.50–1.01) (Cardwell *et al*, 2012).

Lung cancer. Our study showed no association between lung cancer risk and bisphosphonate use. This is consistent with the only other epidemiological study (Cardwell *et al*, 2012), which also found a reduced risk for long-term users (AHR: 0.86, 95% CI: 0.63–1.17) but it was not statistically significant. The antitumour properties of bisphosphonates might not be applicable in the lung because it is not a site where bisphosphonates accumulate (Fournier *et al*, 2002). The sensitivity analysis for all prescriptions including the past 6 months showed a significantly increased risk but without any duration relationship and in particular for short-term users. This may simply suggest that a common symptom of lung cancer, bone pain, is sometimes initially mistaken as osteoporosis.

Pancreas. Although none of the database analyses showed statistically significantly decreased risks associated with bisphosphonate use, the combined analysis did. The findings were consistent with the only other epidemiological study (Cardwell *et al*, 2012) for this cancer site (AHR: 0.84, 95% CI: 0.53–1.35). The effects of bisphosphonates on growth and apoptosis in

Table 4. Bisphosphonate use for individual drugs in cancer cases and controls, numbers and odds ratios (95% CIs) compared with non-use by database

	QResearch			CPRD			Combined analysis	
	Cases/ controls	Adjusted*odds ratio (95% CI)	*P*-value	Cases/ controls	Adjusted*odds ratio (95% CI)	*P*-value	Pooled odds ratio (95% CI)	*P*-value
Breast, alendronate	Trend		**0.7**	Trend		**0.04**	Trend	**0.3**
Any use	901/4627	0.98 (0.89–1.07)	0.6	892/4648	0.96 (0.88–1.05)	0.3	0.97 (0.91–1.03)	0.3
Short term (less than a year)	409/2148	0.95 (0.85–1.08)	0.4	414/2085	1.00 (0.89–1.13)	0.9	0.98 (0.90–1.06)	0.6
Long term (at least a year)	492/2479	1.00 (0.89–1.12)	1.0	478/2563	0.92 (0.83–1.03)	0.2	0.96 (0.89–1.04)	0.3
Breast, etidronate	Trend		**0.1**	Trend		**0.8**	Trend	**0.2**
Any use	453/2437	0.90 (0.81–1.01)	0.07	412/2246	0.91 (0.81–1.02)	0.1	0.91 (0.84–0.98)	0.02
Short term (less than a year)	176/893	0.95 (0.80–1.13)	0.6	179/1018	0.88 (0.74–1.03)	0.1	0.91 (0.81–1.03)	0.1
Long term (at least a year)	277/1544	0.87 (0.76–1.00)	0.05	233/1228	0.94 (0.81–1.09)	0.4	0.90 (0.82–1.00)	0.04
Breast, risedronate	Trend		**0.9**	Trend		**0.3**	Trend	**0.5**
Any use	281/1537	0.92 (0.80–1.06)	0.2	279/1492	0.96 (0.84–1.11)	0.6	0.94 (0.85–1.04)	0.2
Short term (less than a year)	139/803	0.89 (0.73–1.07)	0.2	118/733	0.83 (0.68–1.02)	0.07	0.86 (0.75–0.99)	0.03
Long term (at least a year)	142/734	0.96 (0.80–1.16)	0.7	161/759	1.08 (0.90–1.29)	0.4	1.02 (0.90–1.17)	0.7
Prostate, alendronate	Trend		**0.009**	Trend		**0.05**	Trend	**0.001**
Any use	292/1506	0.81 (0.70–0.93)	0.004	257/1299	0.88 (0.75–1.03)	0.1	0.84 (0.75–0.93)	0.001
Short term (less than a year)	142/692	0.86 (0.71–1.05)	0.1	120/591	0.91 (0.74–1.13)	0.4	0.89 (0.77–1.02)	0.1
Long term (at least a year)	150/814	0.75 (0.62–0.92)	0.004	137/708	0.86 (0.70–1.05)	0.1	0.80 (0.70–0.92)	0.002
Prostate, etidronate	Trend		**0.9**	Trend		**0.6**	Trend	**0.8**
any use	141/567	1.06 (0.87–1.30)	0.5	104/496	0.88 (0.70–1.10)	0.3	0.98 (0.84–1.14)	0.8
Short term (less than a year)	55/223	1.10 (0.81–1.48)	0.6	34/209	0.69 (0.47–1.00)	0.05	0.91 (0.72–1.15)	0.4
Long term (at least a year)	86/344	1.04 (0.81–1.33)	0.8	70/287	1.02 (0.77–1.34)	0.9	1.03 (0.86–1.24)	0.8
Prostate, risedronate	Trend		**0.9**	Trend		**0.3**	Trend	**0.4**
Any use	99/455	0.98 (0.77–1.23)	0.8	70/374	0.87 (0.66–1.13)	0.3	0.93 (0.78–1.10)	0.4
Short term (less than a year)	38/208	0.83 (0.58–1.18)	0.3	32/165	0.90 (0.61–1.33)	0.6	0.86 (0.66–1.12)	0.3
Long term (at least a year)	61/247	1.09 (0.81–1.46)	0.6	38/209	0.83 (0.58–1.19)	0.3	0.98 (0.78–1.23)	0.8
Lung, alendronate	Trend		**0.5**	Trend		**0.4**	Trend	**0.3**
Any use	696/2542	1.00 (0.89–1.12)	0.9	756/2701	1.05 (0.94–1.18)	0.4	1.03 (0.95–1.11)	0.5
Short term (less than a year)	339/1114	1.07 (0.92–1.24)	0.4	386/1220	1.16 (1.01–1.34)	0.04	1.12 (1.01–1.24)	0.04
Long term (at least a year)	357/1428	0.93 (0.80–1.08)	0.3	370/1481	0.96 (0.83–1.11)	0.6	0.95 (0.85–1.05)	0.3
Lung, etidronate	Trend		**0.8**	Trend		**0.1**	Trend	**0.2**
Any use	338/1262	1.04 (0.90–1.20)	0.6	356/1200	1.13 (0.98–1.31)	0.09	1.09 (0.98–1.20)	0.2
Short term (less than a year)	139/500	1.02 (0.82–1.27)	0.9	161/502	1.15 (0.94–1.42)	0.2	1.09 (0.94–1.27)	0.3
Long term (at least a year)	199/762	1.05 (0.87–1.26)	0.6	195/698	1.12 (0.93–1.35)	0.2	1.08 (0.95–1.24)	0.2
Lung, risedronate			**0.9**	Trend		**0.7**	Trend	**0.7**
Any use	261/848	1.02 (0.86–1.21)	0.8	252/829	1.11 (0.94–1.31)	0.2	1.06 (0.94–1.20)	0.3
Short term (less than a year)	125/380	1.05 (0.83–1.33)	0.7	126/412	1.06 (0.85–1.34)	0.6	1.06 (0.90–1.25)	0.5
Long term (at least a year)	136/468	0.98 (0.78–1.22)	0.8	126/417	1.13 (0.90–1.42)	0.3	1.05 (0.89–1.23)	0.6

Abbreviations: CI = confidence interval; CPRD = Clinical Practice Research Datalink. *Adjusted for the confounders listed in the footnote for Table 2. Bold values indicate statistical significance.

pancreatic cancer cells have been shown *in vitro* (Tassone *et al*, 2003) and *in vivo* (Takiguchi *et al*, 2012). No duration relationship suggests, however, that as with breast and prostate cancers bisphosphonates might inhibit growth only of already existing tumours.

Other cancers. Our study has not shown any significant associations between overall use of bisphosphonates and risk of endometrial, ovarian or cervical cancers, but the numbers for endometrial and cervical cancers were very low and the findings were not consistent across the databases. An association found between ovarian cancer and risedronate use was found only in QResearch data and without a duration relationship. Previous much smaller studies have suggested a possible decreased risk for bisphosphonate users for ovarian (AHR: 0.64, 95% CI: 0.40–1.03) and endometrial cancers (AHR: 0.64, 95% CI: 0.38–1.08 (Cardwell *et al*, 2012) and AOR: 0.7, 95% CI: 0.4–1.2 (Fortuny *et al*, 2009)), but the results also did not reach a statistically significant level. Our finding of a decreased risk for bladder cancer, which was not however statistically significant, was consistent with a previous study (AHR: 0.67, 95% CI: 0.41–1.10, $P = 0.11$) (Cardwell *et al*, 2012). No statistically significant associations between bisphosphonate use and melanoma risk were found, similar to the earlier study (AHR: 0.80, 95% CI: 0.53–1.20) (Cardwell *et al*, 2012).

CONCLUSION

We have conducted a series of large population-based case–control studies in two primary-care databases examining the association of bisphosphonates with risks of common cancers in the general population and found associations with reduced risks for breast, prostate and pancreatic cancers, but with no duration relationship and only in short-term users. A duration-dependent reduced risk associated with alendronate use was found for prostate cancer, but only in the QResearch data. This study does not provide enough evidence to conclude that bisphosphonates have protective effects on cancer, but the results are reassuring regarding the safety of bisphosphonates.

ACKNOWLEDGEMENTS

We acknowledge the contribution of EMIS and the University of Nottingham for expertise in creating and maintaining QResearch and to the EMIS practices, which contribute data without whom this research would not be possible.

REFERENCES

Bangalore S, Kumar S, Kjeldsen SE, Makani H, Grossman E, Wetterslev J, Gupta AK, Sever PS, Gluud C, Messerli FH (2011) Antihypertensive drugs and risk of cancer: network meta-analyses and trial sequential analyses of 324,168 participants from randomised trials. *Lancet Oncol* **12**(1): 65–82.

British Medical Association and Royal Pharmaceutical Society of Great Britain (2008) *British National Formulary* Vol. 56. British Medical Association and Royal Pharmaceutical Society of Great Britain: London, UK.

Burbos N, Musonda P, Giarenis I, Shiner AM, Giamougiannis P, Morris EP, Nieto JJ (2010) Predicting the risk of endometrial cancer in postmenopausal women presenting with vaginal bleeding: the Norwich DEFAB risk assessment tool. *Br J Cancer* **102**(8): 1201–1206.

Cardwell CR, Abnet CC, Veal P, Hughes CM, Cantwell MM, Murray LJ (2012) Exposure to oral bisphosphonates and risk of cancer. *Int J Cancer* **131**(5): E717–E725.

Chang ET, Smedby KE, Hjalgrim H, Glimelius B, Adami H-O (2006) Reliability of self-reported family history of cancer in a large case–control study of lymphoma. *J Natl Cancer Inst* **98**(1): 61–68.

Chen Z, Arendell L, Aickin M, Cauley J, Lewis CE, Chlebowski R (2008) Hip bone density predicts breast cancer risk independently of Gail score. *Cancer* **113**(5): 907–915.

Chlebowski RT, Chen Z, Cauley JA, Anderson G, Rodabough RJ, McTiernan A, Lane DS, Manson JE, Snetselaar L, Yasmeen S, O'Sullivan MJ, Safford M, Hendrix SL, Wallace RB (2010) Oral bisphosphonate use and breast cancer incidence in postmenopausal women. *J Clin Oncol* **28**(22): 3582–3590.

Corley DA, Kubo A, Zhao W, Quesenberry C (2010) Proton pump inhibitors and histamine-2 receptor antagonists are associated with hip fractures among at-risk patients. *Gastroenterology* **139**(1): 93–101.

Coussens LM, Werb Z (2002) Inflammation and cancer. *Nature* **420**(6917): 860–867.

Croucher P, Jagdev S, Coleman R (2003) The anti-tumor potential of zoledronic acid. *Breast* **12**(Suppl 2): S30–S36.

Ferlay J, Shin H-R, Bray F, Forman D, Mathers C, Parkin DM (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* **127**(12): 2893–2917.

Fortuny J, Sima C, Bayuga S, Wilcox H, Pulick K, Faulkner S, Zauber AG, Olson SH (2009) Risk of endometrial cancer in relation to medical conditions and medication use. *Cancer Epidemiol Biomarkers Prev* **18**(5): 1448–1456.

Fournier P, Boissier S, Filleur S, Guglielmi J, Cabon F, Colombel M, Clézardin P (2002) Bisphosphonates inhibit angiogenesis *in vitro* and testosterone-stimulated vascular regrowth in the ventral prostate in castrated rats. *Cancer Res* **62**(22): 6538–6544.

Gnant M (2010) Can oral bisphosphonates really reduce the risk of breast cancer in healthy women? *J Clin Oncol* **28**(22): 3548–3551.

Gonzalez-Perez A, Garcia Rodriguez L, Lopez-Ridaura R (2003) Effects of non-steroidal anti-inflammatory drugs on cancer sites other than the colon and rectum: a meta-analysis. *BMC Cancer* **3**(1): 28.

Green J, Czanner G, Reeves G, Watson J, Wise L, Beral V (2010) Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case–control analysis within a UK primary care cohort. *BMJ* **341**: c4444.

Guise TA (2008) Antitumor effects of bisphosphonates: promising preclinical evidence. *Cancer Treat Rev* **34**(Suppl 1): S19–S24.

Hecht SS (2008) Etiology of cancer: tobacco. In *DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology,* Vol. 1, Chapter 11 pp 147–155. Wolters Kluwer/Lippincott Williams & Wilkins: Philadelphia, PA, USA.

Henderson KD, Bernstein L (2008) Etiology of cancer: obesity and physical activity. In *DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology,* Vol. 1, Chapter 18pp 239–244. Wolters Kluwer/Lippincott Williams & Wilkins: Philadelphia, PA, USA.

Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ (2010) Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* **69**(1): 4–14.

Hippisley-Cox J, Coupland C (2009) Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. *BMJ* **339**: b4229.

Hippisley-Cox J, Coupland C (2010) Unintended effects of statins in men and women in England and Wales: population based cohort study using the QResearch database. *BMJ* **340**: c2197.

Hippisley-Cox J, Hammersley V, Pringle M, Coupland C, Crown N, Wright L (2004) How useful are General Practice databases for research? Analysis of their accuracy and completeness in one research network. *Health Inform J* **10**: 91–109.

Jick SS, Kaye JA, Vasilakis-Scaramozza C, Rodríguez LAG, Ruigómez A, Meier CR, Schlienger RG, Black C, Jick H (2003) Validity of the General Practice Research Database. *J Hum Pharmacol Drug Ther* **23**(5): 686–689.

Kanis J, McCloskey E, Johansson H, Strom O, Borgstrom F, Oden A (2008) Case finding for the management of osteoporosis with FRAX® – assessment and intervention thresholds for the UK. *Osteoporos Int* **19**: 1395–1408.

Mai PL, Garceau AO, Graubard BI, Dunn M, McNeel TS, Gonsalves L, Gail MH, Greene MH, Willis GB, Wideroff L (2011) Confirmation of family cancer history reported in a population-based survey. *J Natl Cancer Inst* **103**(10): 788–797.

Mauck KF, Clarke BL (2006) Diagnosis, screening, prevention, and treatment of osteoporosis. *Mayo Clin Proc* **81**(5): 662–672.

McGlynn KA, Gridley G, Mellemkjaer L, Brinton LA, Anderson KC, Caporaso NE, Landgren O, Olsen JH (2008) Risks of cancer among a cohort of 23,935 men and women with osteoporosis. *Int J Cancer* **122**(8): 1879–1884.

Mocellin S (2011) Vitamin D and cancer: deciphering the truth. *Biochim Biophys Acta* **1816**(2): 172–178.

National Institute for Health and Clinical Excellence (2008a) Alendronate, etidronate, risedronate, strontium ranelate and raloxifen for preventing bone fractures in postmenopausal women with osteoporosis who have not had a fracture. *Information About NICE Technology Appraisal Guidance*. p 160.

National Institute for Health and Clinical Excellence (2008b) Alendronate, etidronate, risedronate, strontium ranelate and teriparatide for preventing bone fractures in postmenopausal women with osteoporosis who have already had a fracture. *Information About NICE Technology Appraisal Guidance*. p 161.

Newcomb PA, Trentham-Dietz A, Hampton JM (2010) Bisphosphonates for osteoporosis treatment are associated with reduced breast cancer risk. *Br J Cancer* **102**(5): 799–802.

Parker C, Coupland C, Hippisley-Cox J (2010) Antipsychotic drugs and risk of venous thromboembolism: nested case–control study. *BMJ* **341**: c4245.

Rennert G, Pinchev M, Rennert HS (2010) Use of bisphosphonates and risk of postmenopausal breast cancer. *J Clin Oncol* **28**(22): 3577–3581.

Royston P (2004) Multiple imputation of missing values. *Stata J* **4**(3): 227–241.

Royston P (2005) Multiple imputation of missing values: update of ice. *Stata J* **5**(4): 527–536.

Schütze M, Boeing H, Pischon T, Rehm J, Kehoe T, Gmel G, Olsen A, Tjønneland AM, Dahm CC, Overvad K, Clavel-Chapelon F, Boutron-Ruault M-C, Trichopoulou A, Benetou V, Zylis D, Kaaks R, Rohrmann S, Palli D, Berrino F, Tumino R, Vineis P, Rodríguez L, Agudo A, Sánchez M-J, Dorronsoro M, Chirlaque M-D, Barricarte A, Peeters PH, van Gils CH, Khaw K-T, Wareham N, Allen NE, Key TJ, Boffetta P, Slimani N, Jenab M, Romaguera D, Wark PA, Riboli E, Bergmann MM (2011) Alcohol attributable burden of incidence of cancer in eight European countries based on results from prospective cohort study. *BMJ* **342**: d1584.

Takiguchi S, Nishino Y, Inoue K, Ikeda M, Kataoka Y, Matsusue K, Nishiyama K, Iguchi H (2012) The bisphosphonate incadronate inhibits intraperitoneal dissemination in an in vivo pancreatic cancer model. *Oncol Rep* **28**(1): 111–116.

Tassone P, Tagliaferri P, Viscomi C, Palmieri C, Caraglia M, D'Alessandro A, Galea A, Goel A, Abbruzzese A, Boland CR, Venuta S (2003) Zoledronic acid induces antiproliferative and apoptotic effects in human pancreatic cancer cells *in vitro*. *Br J Cancer* **88**(12): 1971–1978.

Thomas E, Brewster DH, Black RJ, Macfarlane GJ (2000) Risk of malignancy among patients with rheumatic conditions. *Int J Cancer* **88**(3): 497–502.

Tuomela J, Valta M, Vaananen K, Harkonen P (2008) Alendronate decreases orthotopic PC-3 prostate tumor growth and metastasis to prostate-draining lymph nodes in nude mice. *BMC Cancer* **8**(1): 81.

Vestergaard P, Fischer L, Mele M, Mosekilde L, Christiansen P (2011) Use of bisphosphonates and risk of breast cancer. *Calcif Tissue Int* **88**(4): 255–262.

Vincent A, Herman J, Schulick R, Hruban RH, Goggins M (2011) Pancreatic cancer. *Lancet* **378**(9791): 607–620.

Vinogradova Y, Coupland C, Hippisley-Cox J (2012) Exposure to bisphosphonates and risk of cancer: a protocol for nested case–control studies using the QResearch primary care database. *BMJ Open* **2**(1): e000548.

Walker AJ, Card T, Bates TE, Muir K (2011) Tricyclic antidepressants and the incidence of certain cancers: a study using the GPRD. *Br J Cancer* **104**(1): 193–197.

Watts NB, Diab DL (2010) Long-term use of bisphosphonates in osteoporosis. *J Clin Endocrinol Metab* **95**(4): 1555–1565.

Paper 7

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to combined oral contraceptives and risk of venous thromboembolism: a protocol for nested case-control studies using the QResearch and the CPRD databases. *BMJ Open* 2014;4(4):e004499.

BMJ Open

Exposure to combined oral contraceptives and risk of venous thromboembolism: a protocol for nested case–control studies using the QResearch and the CPRD databases

Yana Vinogradova, Carol Coupland, Julia Hippisley-Cox

To cite: Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to combined oral contraceptives and risk of venous thromboembolism: a protocol for nested case–control studies using the QResearch and the CPRD databases. *BMJ Open* 2014;**4**:e004499. doi:10.1136/bmjopen-2013-004499

▸ Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2013-004499).

Received 18 November 2013
Revised 5 March 2014
Accepted 7 March 2014

Division of Primary Care, University of Nottingham, University Park, Nottingham, UK

Correspondence to
Yana Vinogradova; Yana.Vinogradova@nottingham.ac.uk

ABSTRACT

Introduction: Many studies have found an increased risk of venous thromboembolism (VTE) associated with the use of combined hormonal contraceptives, but various methodologies have been used in the study design relating to definition of VTE event and the selection of appropriate cases for analysis. This study will focus on common oral hormonal contraceptives, including compositions with cyproterone because of their contraceptive effect and will perform a number of sensitivity analyses to compare findings with previous studies.

Methods and analysis: 2 nested case–control studies will be based on the general population using records from UK general practices within the QResearch and Clinical Practice Research Datalink databases. Cases will be female patients aged 15–49 with primary VTE diagnosed between 2001 and 2013. Each case will be matched by age, year of birth and practice to five female controls, who are alive and registered with the practice at the time of diagnosis of the case (index date). Exposure to different hormonal contraceptives will be defined as at least one prescription for that contraceptive in the year before the index date. The effects of duration and the length of any gap since last use will also be investigated. Conditional logistic regression will be applied to calculate ORs adjusted for smoking, ethnicity, comorbidities and use of other medications. Possible indications for prescribing hormonal contraceptives, such as menstrual disorders, acne or hirsutism will be included in the analyses as confounding factors. A number of sensitivity analyses will be carried out.

Ethics and dissemination: The initial protocol has been reviewed and approved by ISAC (Independent Scientific Advisory Committee) for Medicine and Healthcare Products Regulatory Agency Database Research. The project has also been reviewed by QResearch and meets the requirements of the Trent Research Ethics Committee. The results will be published in a peer-reviewed journal.

Strengths and limitations of this study

- Primary care research databases.
- Large size and great statistical power.
- A range of sensitivity analyses to compare the results with other studies.
- Results being adjusted for all confounders for which data are available.
- Prescription-based study.
- Possible uncertainty in the diagnosis of venous thromboembolism.
- Underestimation of hormonal contraceptive use.
- Lack of information on some confounding factors that might affect the choice of contraceptive drug.

INTRODUCTION

An increased risk of thrombosis in users of hormonal contraceptives has been identified by a number of studies, and this has resulted in British National Formulary (BNF) recommendations[1] to consider risk factors for venous thromboembolism (VTE) before prescribing the drugs and to avoid using them if two or more risk factors are present. Since the onset of oral contraceptive use in the general female population in the 1960s, studies have demonstrated associations between the drugs and a range of adverse side effects, including an increased risk of VTE. The composition of hormonal contraceptives has, therefore, changed over time. A 'second generation' aimed to reduce the increased VTE risk, lowering the oestrogen component by using potent testosterone-derived progestogens.[2] A later 'third-generation' was introduced to lower the androgenic and vascular risk by introducing progestogens with low androgenic activity[3] and to reduce arterial vascular impact.[2] Effects on VTE from third-generation

contraceptive use, have, however, been complex, with some increased risks reported.[4]

There are a large number of observational studies looking at the effect of contraceptive drugs on the general female population, but there are three key methodological issues which have been handled very differently across these. The first concerns verification of the VTE diagnosis. Standardised criteria for diagnostic categories include four levels of verification: positive imaging tests (eg, positive Doppler ultrasound or impedance plethysmography) and subsequent therapy (1: definite VTE), uncertain imaging tests with subsequent therapy (2: probable VTE), positive imaging tests without subsequent therapy or negative imaging tests but with subsequent therapy (3: possible VTE), and 'typical symptoms' without confirming tests or therapy (4: potential VTE).[5]

To date, observational studies have treated the verification of VTE in a number of different ways. An Austrian study distinguished between confirmed and not confirmed cases, concentrating on cases with definite and probable VTE for the main analysis and performing additional analysis on the sample including possible and potential VTE cases, which produced statistically identical results to their main analysis.[6] An Israeli study based on a healthcare provider's database used clinical records only without any verification.[7] A Danish study based on national healthcare databases used anticoagulation prescriptions for verification and produced a stratified analysis of confirmed and non-confirmed diagnoses demonstrating a twofold to threefold higher risk associated with VTE in the confirmed group.[8] A number of studies based on electronically collected data included cases with diagnosis of VTE confirmed with subsequent anticoagulant prescriptions but without using any diagnostic tests.[4 9–12] A Dutch study based on hospital and general practitioners' (GP) records required confirmation of VTE diagnosis with Doppler ultrasonography.[13] These variations in levels of verification and differences in analysis strategy both complicate comparisons of study findings.

The second area of variation between studies lies in sample selection. For example, although women with oophorectomy, hysterectomy or sterilisation should not remain in the group potentially exposed to contraceptives, of the major studies with reference to the no use group only Lidegaard *et al*[8] mention exclusion of such patients. As important is the difference in handling of non-idiopathic cases—those with other potentially important proximate causes and risk factors. Almost all studies were aligned in excluding women with previous VTE (as the studies were focused on incident cases) and the majority[8–10 13 14] in excluding pregnant and postpartum women (unlikely users of contraceptive drugs and with a higher risk of VTE), but the handling of non-idiopathic cases based on morbidities has varied significantly.

The study of Farmer *et al*[4] excluded patients with recent major surgery and trauma, cancer and congenital heart disease, while the studies of Jick and coauthors[9 11] added renal failure, chronic cardiovascular disease, inflammatory or autoimmune conditions and an operation or major trauma before the diagnosis as exclusion criteria. The study of Parkin *et al*[10] extended the exclusion list even further with diabetes type I, colitis, systemic lupus erythematosus, spondylitis, cystic fibrosis, psoriatic arthritis and coagulation disturbances. The study of Lidegaard *et al*,[8] however, excluded only those with selected cancers and coagulation disturbances, while a number of studies[15–18] did not exclude any such morbidity-related idiopathic cases. A study by Heinemann *et al*[6] identified a subgroup of idiopathic cases and used additional analysis to demonstrate that the ORs for the selected group were twice those for the whole study sample.

The third methodological issue involves the related issues of exposure definition and reference group selection. Some studies were based only on current users, estimating the risk associated with use of one drug in comparison with another,[7 9 10 15 16] while others compared current users with the non-exposed group.[6 8 18] 'Current use' has also had a range of definitions, which is problematic because the increased risk of VTE in patients on oral contraceptives decreases after therapy stops and disappears within 3 months.[19] The study of Heinemann *et al*[6] considered a patient as a current user for 6 weeks after discontinuation, while the studies of Jick *et al*[9] and Parkin *et al*[10] extended the period of current use for only 30 days. The study of Lidegaard *et al*[8] allowed 4 weeks after the end date of the prescription before changing a woman's exposure status to previous user, while the study of Gronich *et al*[7] allowed 3 months. Seeger *et al*[15] considered women as current users only if they had not reached the end date for the prescription, while in the questionnaire-based studies[16 18] current use was derived from the responses of participants.

Twenty-six studies based on data up to 2013 contributed to a recent meta-analysis[20] for combined oral contraceptives, which demonstrated a twofold or more increased risk of VTE in users of any type or generation of 'the pill' compared with non-users. All the studies listed above were included in the meta-analysis but had a wide variation in estimates because of their heterogeneity of approach, in particular with respect to the definition of cases, inclusion criteria and reference group selection and definition of exposure. This overview shows that there are no established criteria for selecting patients for assessment of the VTE risk associated with use of oral contraceptives. Excluding cases without anticoagulant therapy might introduce a selection bias as doctors may be more likely to start medical treatment in patients on contraceptive drugs even with mild symptoms of VTE.[21] Excluding non-idiopathic cases restricts analysis to relatively healthy patients but does not remove patients with known risk factors such as smoking, obesity or other unmeasured lifestyle factors.

Established risk factors, however, do not prevent doctors from prescribing contraceptive drugs and their decisions are affected by all information available to them, so the question of how inclusive or exclusive the sample should be is important from a practical point of view.

The proposed nested case–control studies based on the general female population will investigate the association between the use of most common hormonal contraceptives and risk of VTE adjusted for indications other than contraception (polycystic ovarian syndrome (PCOS) and menstrual disorders), comorbidities affecting prescribing and other confounding factors. In terms of exclusions and inclusions, the study will perform a number of sensitivity analyses to make it comparable with other studies. It will provide an overview of the risks associated with the currently most common types of oral contraceptives and increase its power by combining the results obtained from the two largest electronic medical records databases, QResearch and Clinical Practice Research Datalink (CPRD).

METHODS AND ANALYSIS

Data source

This study will use two separate data sources—the QResearch primary care research database (http://www.qresearch.org and the CPRD, http://www.cprd.com). Both consist of routinely collected data from GP clinical computer systems and each contains information from around 7% of all UK general practices. The information recorded on the databases includes patient demographics (year of birth and sex), characteristics (height, weight and smoking status), clinical diagnoses, symptoms and prescribed medications, including repeat prescriptions. Both databases have been created for research purposes and linked to other sources of information, such as Hospital Episode Statistics (HES) and Office of National Statistics mortality data. Both have been validated using other sources of information in the UK, which has demonstrated their accuracy and completeness.[22] Although QResearch has not been validated as extensively as CPRD, a recent study on risk of cancer and use of bisphosphonates based on these databases demonstrated similar prevalence in outcomes and prescribing.[23 24] Both databases have been used in previous studies of VTE associated with prescription information.[10 25]

Sample selection

The studies will use records from UK general practices within the QResearch database and within the CPRD database. An open cohort of women from each database will be identified, all between the age of 15 and 49 years, registered with the study practices during the study period between 1 January 2001 and 31 December 2013. The right censor date will be the earliest of the following where applicable: date of diagnosis of VTE, date of death, date of leaving the practice, date of the latest download of data and study end date. Diagnosis of VTE will be based on recording in the electronic patient records using READ codes—the list of READ codes is presented in table 1.

Within each of these two cohorts we will design two nested case–control studies with incident cases of VTE registered during the study period. Cases will be individually matched with up to five female controls with the same year of birth, age and from the same general practice. The controls will be selected using incidence density sampling and allocated an index date, which is the date on which their matched case was first diagnosed with VTE.

Table 1 Read codes for VTE used in QResearch and CPRD data extraction

Read codes	Read description
F051.00	Thrombosis of central nervous system venous sinuses
F051000	Thrombosis cavernous sinus
F051100	Thrombosis of superior longitudinal sinus
F051200	Thrombosis lateral sinus
F051300	Thrombosis transverse sinus
F051z00	Thrombosis of central nervous system venous sinus NOS
F423811	Retinal vein thrombosis
G401	Pulmonary embolism
G401-1	Infarction—pulmonary
G401-2	Pulmonary embolus
G4010	Postoperative pulmonary embolus
G676	Non-pyogenic venous sinus thrombosis
G801-1	DVT
G801-2	DVT, leg
G801-3	DVT
G801-99	DVT—leg
G801C	DVT of leg related to air travel
G801D	Deep vein thrombosis of lower limb
G801D-99	DVT—leg
G801E	DVT of leg related to intravenous drug use
G801F	DVT of peroneal vein
G820	Budd-Chiari syndrome (hepatic vein thrombosis)
G820-1	Hepatic vein thrombosis
G822	Embolism and thrombosis of the vena cava
G823	Embolism and thrombosis of the renal vein
G824.00	Axillary vein thrombosis
G825.00	Thrombosis of subclavian vein
G826.00	Thrombosis of internal jugular vein
G827	Thrombosis of external jugular vein
G82y.00	Other embolism and thrombosis
G82z.00	Embolism and thrombosis NOS
G82z0	Venous embolism NOS
G82z1	Venous thrombosis NOS
G82zz00	Embolism and thrombosis NOS
SP122	Postoperative deep vein thrombosis

CPRD, Clinical Practice Research Datalink; DVT, deep vein thrombosis; NOS, not otherwise specified; VTE, venous thromboembolism.

Exclusions

Cases and controls with any previous VTE diagnosis prior to entry into the study will be excluded. Cases with anticoagulant prescriptions (BNF 2.8) earlier than 6 weeks prior to the diagnosis will be excluded since this could indicate a previous VTE, and controls with such prescriptions before the index date will also be excluded.

Cases and controls will be excluded from the analysis if they have conditions preventing use of contraceptives such as oophorectomy, hysterectomy and sterilisation. We will also exclude women pregnant at the index date or in the first 3 months after delivery (using pregnancy codes and an estimated conception date as delivery date minus 280 days or delivery date minus gestational age if recorded), because these patients have a higher VTE risk[26] and because it is less likely for breast feeding women to have been using hormonal contraceptives.

Eligible cases and controls will have at least 1 year of records prior to the index date.

Exposure

The observational period for assessing exposure for each patient will be defined as the last year before the index date.

Exposure to hormonal contraceptives will be based on all prescriptions for combined hormonal and progestogen-only contraceptives within the 1 year observation period (BNF 7.3.1, 7.3.2) and hormonal treatment of acne (co-cyprindiol or cyproterone, from BNF 13.6.2). Cyproterone will be included as a hormonal contraceptive because it has a similar effect to progestogen on the release of testosterone by ovaries.[27] A participant will be considered as ever exposed if they had at least one prescription for a hormonal contraceptive.

The main focus will be on combined oral contraceptives. We will consider the compositions containing levonorgestrel, desogestrel, norgestimate, norethisterone, gestodene, drospirenone and cyproterone. As an association of increased risk of VTE in transdermal versus oral contraceptive users has been found,[28] women exposed to non-oral combined contraceptives will be identified and kept in the analysis. Progestogen-only drugs are not expected to be associated with an increased risk of VTE[8] but will be kept in the analysis for comparison purposes, with oral and non-oral preparations as two different types of exposure. For all analyses non-users of hormonal contraceptives in the previous year will be the reference category.

Numbers permitting, dosages of oestrogen, 20 or 30 mg and more of ethinylestradiol, will be analysed for the most common compositions: norethisterone, desogestrel and gestodene.

The duration of exposure will be assessed by calculating the number of days prescribed within the previous year. If the gap between the end of one prescription and the start of the next is not more than 30 days, use will be considered as continuous and the duration of the prescriptions will be summed. If a gap between prescriptions is more than 30 days only the latest period of exposure will be considered.

Recency of use will be analysed by calculating the gap in days between the estimated date for the last use and the index date, and categorising it as: current use (using drugs at the index date or the last use was no more than 28 days before the index date), past use (last use between 29 and 365 days before the index date) and no use in last year. If a woman was exposed to more than one oral contraceptive in the 28 days before the index date, only the latest received drug will be analysed, and a variable indicating whether or not women had switched in the last 28 days will be included in the analysis.

We will estimate the effect of the duration of the last exposure by categorising it as up to 84 days (short term) and more than 84 days (long term). The cut point of 84 days (12 weeks) is chosen because VTE risk decreases after 3 months of exposure[19] and 84 days is the most common length of a contraceptive prescription. We will combine recency and duration of exposure to give four categories for each drug exposure: current use with short-term exposure, current use with long-term exposure, past exposure and no use in the last year.

Confounding factors

All analyses will be adjusted for confounders established as risk factors for VTE because they are listed in National Health Service (NHS) guidelines[29] and affect doctors' decisions about prescribing hormonal contraceptives. The list will include comorbidities associated with increased risk of VTE[30]: cancer, congestive cardiac failure, varicose veins, cardiovascular disease, rheumatoid arthritis, systemic lupus erythematosus, chronic renal disease, asthma, chronic obstructive pulmonary disease, Crohn's or ulcerative colitis and coagulation disturbances (Leiden factor V, protein C and S deficiencies).[31] Particular medical events will also be included if recorded in the past 6 months prior to the index date: acute infections, surgery, hospitalisation, leg or hip fracture.[12 30] Patients with these comorbidities and conditions will be identified as non-idiopathic cases for further sensitivity analysis.

Other confounders—patients' characteristics measured at the closest date before the index date—will be: body mass index (BMI, continuous variable)[7], smoking status (current smoker: light 1–9 cigarettes/day, medium 10–19, heavy 20 or more, ex-smoker and non-smoker)[32], alcohol consumption and ethnicity (White, Black, Asian, other).[33]

As there is likely to be a large group of women taking hormonal contraceptives for treatment of PCOS, this condition will also be included because of associations with increased risk of VTE.[34] Other reasons for combined hormonal contraceptive use, such as acne, hirsutism and menstrual disorders, will be included into

analysis if the OR for at least one of the exposure variables is changed by more than 10%.

Statistical analysis

Conditional logistic regression will be used to estimate ORs with 95% CIs for VTE. The initial analysis model will determine the unadjusted ORs for VTE associated with the key exposure variables of interest (specific types of drugs, recency of use and duration and dose). A multivariable model will determine the OR for VTE associated with hormonal contraceptive prescriptions, adjusted for the confounding variables listed above. The main analyses will be run on all cases with VTE identified from the general practice data and their matched controls. A sensitivity analysis will be run on the subgroup of cases and their matched controls where the case diagnosis is supported by thrombolytic prescriptions in the 6 weeks before or after the VTE diagnosis. A second sensitivity analysis will be run on idiopathic cases and their controls, excluding from the analysis all cases and controls with medical conditions and recent events established as VTE risk factors. A third sensitivity analysis will be run on all non-idiopathic cases and controls.

For practices linked to HES data another sensitivity analysis will be run. New cases of VTE identified in HES will be added to the analysis and controls with VTE recorded in HES prior to the index date will be removed.

As the proportion of women using contraceptive clinics (where the data on contraception is not recorded in their GP records) is higher (10%) for a younger group (15–24 years) compared with 3% for women 25 years and older,[35] separate subgroup analyses will be run on the older and younger group, and we will carry out a test for interaction with age.

As BMI, smoking status and alcohol consumption may be important confounders but have non-negligible numbers of missing data, multiple imputation will be used to impute missing values.[36] Ten imputed datasets will be created. Index year, case/control status, age, years of records, potential confounders and exposure to hormonal contraceptives and other drugs, will be included in the imputation model. The distribution for BMI will be assessed and, if not normal, a transformation will be carried out prior to inclusion in the imputation model. Characteristics of women with missing values and with complete data will be compared to assess whether it is plausible that data are missing at random. A sensitivity analysis restricted to women without missing data for BMI, smoking status and alcohol consumption will also be performed.

The nested case–control studies for QResearch and CPRD will be carried out separately and in exactly the same way, selecting the same confounders and running the same procedures. All observations will be from general practices in the UK, from the same time period, with similar exposures and using similar methods for recording outcomes. The sizes of the studies are also expected to be similar. Any differences in associations observed across the databases are likely to be caused only by sampling variation. The results from the two studies will then be combined using the method of Mantel and Haenszel for fixed effect models.

A 1% level of statistical significance will be used to allow for multiple comparisons. Stata V.12 will be used for all the analyses.

Sample size calculation

All eligible cases from QResearch and CPRD will be used. According to the Office for National Statistics, combined contraceptives are used by 25% of women aged 15–49 in the UK.[37] For an individual drug with exposure of 5%, 2115 cases and 10 575 controls will be needed to detect a clinically important OR of 1.5 at a significance level of 1% with 90% power. For rarer compositions such as drospirenone or cyproterone with exposure of 1% and a clinically important OR of 2.0, 2882 cases and 14 410 controls will be needed. The numbers of cases from QResearch and CPRD are expected to be fairly similar. The January 2014 version of CPRD contains 8673 cases with first time VTE recorded between 2001 and 2013. After removing pregnant and postpartum women, cases with previous anticoagulant prescriptions and cases with less than a year of medical records, a sample of 5920 cases will be available for analysis in CPRD.

DISCUSSION

This is an observational study based on routinely collected data from two large primary care research databases and will have the strengths and limitations common to all such studies. By combining results from two databases, the study will have greater statistical power than previous studies. It will allow analyses to be carried out investigating the effects of the recency and duration of use for the most commonly used hormonal contraceptives. Because the data on prescriptions and potential confounding variables are routinely and prospectively collected and recorded before the index date, the study will be free from recall bias. Similarly, as all eligible cases and randomly selected controls will be included, there should be no selection bias.

The study will conduct a number of sensitivity analyses to address conflicting methodological issues giving the reader an opportunity to decide which estimates are the most valid.

The limitations of the study will include possible uncertainty in VTE diagnosis. A systematic review based on General Practice Research Database (GPRD) validation studies has reported that, on average, 85% of diagnoses of circulatory system problems recorded on the GP electronic record were confirmed from other data sources.[38] Lawrenson *et al*[39] looked specifically at VTE validation and found that 84% of the diagnoses were supported by hospitalisation or death certificate. Any

misclassification (assuming it is non-differential between cases and controls) will result in underestimation of associations with hormonal contraceptives, shifting the ORs towards unity. The sensitivity analysis on validated diagnosis of VTE along with descriptive statistics will address issues about differential attention to different types of contraceptives raised in the Danish study.[8]

Another limitation is potential underestimation of hormonal contraceptive use. Apart from the GP, hormonal contraceptives are available from other NHS services such as family planning or contraceptive clinics. According to the Health & Social Care Information Centre,[35] approximately 0.6 million women in England are supplied with hormonal contraceptives from contraceptive clinics, which represents about 5% of the targeted population. Although a survey of contraceptive services use in Britain[40] reported that only 59% of responding women would use general practice and 15% would use contraceptive clinics, the response rate of the survey was only 65% and the actual proportions might be smaller. From currently available CPRD data, overall use of hormonal contraceptives based on GP prescriptions is about 50%, so excess of use from other sources will be considered minor.

Other limitations are also common to any general practice database. Information on certain risk factors such as level of physical activity or use of air travel is not reliably recorded so these factors cannot be included in the analyses. Important confounders such as smoking or BMI have non-negligible amounts of missing data so these will be imputed.

The results of this study will help to establish risks of VTE associated with different oral hormonal contraceptive drugs.

ETHICS AND DISSEMINATION

The project has also been reviewed by QResearch and meets the requirements of the Trent Research Ethics Committee. To guarantee the confidentiality of personal and health information only the authors will have access to the data during the study. It will be possible to access the QResearch and CPRD data after the publication of the results but only on premises of the University of Nottingham according to QResearch standard procedures and the CPRD license. The full protocol and statistical code will be available from the authors after the publication of the results.

Acknowledgements The authors would like to acknowledge the helpful contribution of the reviewers, Professor Susan Jick and Professor Øjvind Lidegaard.

Contributors JH-C had the original idea for this study. CC contributed to the development of the idea and the study design. YV reviewed the literature, contributed to the study design and wrote the draft of the manuscript. JH-C and CC critically reviewed the paper. YV is the guarantor of the study. All authors have approved the submitted version.

Competing interests JHC is a professor of clinical epidemiology at the University of Nottingham and unpaid director of QResearch, a not-for-profit organisation which is a joint partnership between the University of Nottingham and EMIS (commercial IT supplier for 60% of general practices in the UK); JHC is also a paid director of ClinRisk Limited, which produces open and closed source software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help improve patient care.

Ethics approval This protocol has been approved by Independent Scientific Advisory Committee for MHRA Database Research (N 13_118R).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement To guarantee the confidentiality of personal and health information only the authors will have access to the data during the study. It will be possible to access the QResearch and CPRD data after the publication of the results but only on premises of the University of Nottingham according to QResearch standard procedures and CPRD license. The full protocol and statistical code will be available from the authors after the publication of the results.

REFERENCES

1. British National Formulary London: British Medical Association and Royal Pharmaceutical Society of Great Britain, 2013.
2. Ouzounian S, Verstraete L, Chabbert-Buffet N. Third-generation oral contraceptives: future implications of current use. *Expert Rev Obstet Gynecol* 2008;3:189–201.
3. LeBlanc E, Laws A. Benefits and risks of third-generation oral contraceptives. *J Gen Intern Med* 1999;14:625–32.
4. Farmer RDT, Lawrenson RA, Todd JC, *et al*. A comparison of the risks of venous thromboembolic disease in association with different combined oral contraceptives. *Br J Clin Pharmacol* 2000;49:580–90.
5. Heinemann LAJ, Lewis MA, Assmann A, *et al*. Case-control studies on venous thromboembolism: bias due to design? A methodological study on venous thromboembolism and steroid hormone use. *Contraception* 2002;65:207–14.
6. Heinemann LAJ, Dinger JC, Assmann A, *et al*. Use of oral contraceptives containing gestodene and risk of venous thromboembolism: outlook 10 years after the third-generation 'pill scare'. *Contraception* 2010;81:401–7.
7. Gronich N, Lavi I, Rennert G. Higher risk of venous thrombosis associated with drospirenone-containing oral contraceptives: a population-based cohort study. *Can Med Assoc J* 2011;183:E1319–25.
8. Lidegaard Ø, Nielsen LH, Skovlund CW, *et al*. Risk of venous thromboembolism from use of oral contraceptives containing different progestogens and oestrogen doses: Danish cohort study, 2001-9. *BMJ* 2011;343:d6423.
9. Jick SS, Hernandez RK. Risk of non-fatal venous thromboembolism in women using oral contraceptives containing drospirenone compared with women using oral contraceptives containing levonorgestrel: case-control study using United States claims data. *BMJ* 2011;342:d2151.
10. Parkin L, Sharples K, Hernandez RK, *et al*. Risk of venous thromboembolism in users of oral contraceptives containing drospirenone or levonorgestrel: nested case-control study based on UK General Practice Research Database. *BMJ* 2011;342:d2139.
11. Jick SS, Kaye JA, Russmann S, *et al*. Risk of nonfatal venous thromboembolism in women using a contraceptive transdermal patch and oral contraceptives containing norgestimate and 35 μg of ethinyl estradiol. *Contraception* 2006;73:223–8.
12. Lawrenson R, Farmer R. Venous thromboembolism and combined oral contraceptives: does the type of progestogen make a difference? *Contraception* 2000;62(2 Supplement 1):S21–8.
13. van Hylckama Vlieg A, Helmerhorst FM, *et al*. The risk of deep venous thrombosis associated with injectable depot–medroxyprogesterone acetate contraceptives or a levonorgestrel intrauterine device. *Arterioscler Thromb Vasc Biol* 2010;30:2297–300.
14. Farmer RDT, Williams TJ, Simpson EL, *et al*. Effect of 1995 pill scare on rates of venous thromboembolism among women taking combined oral contraceptives: analysis of General Practice Research Database. *BMJ* 2000;321:477–9.
15. Seeger JD, Loughlin J, Eng PM, *et al*. Risk of thromboembolism in women taking ethinylestradiol/drospirenone and other oral contraceptives. *Obstet Gynecol* 2007;110:587–93.

16. Dinger JC, Heinemann LAJ, Kühl-Habich D. The safety of a drospirenone-containing oral contraceptive: final results from the European Active Surveillance study on Oral Contraceptives based on 142,475 women-years of observation. *Contraception* 2007;75:344–54.
17. Seaman HE, de Vries CS, Farmer RDT. Venous thromboembolism associated with cyproterone acetate in combination with ethinyloestradiol (Dianette®): observational studies using the UK General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2004;13:427–36.
18. Vlieg AvH, Helmerhorst FM, Vandenbroucke JP, *et al*. The venous thrombotic risk of oral contraceptives, effects of oestrogen dose and progestogen type: results of the MEGA case-control study. *BMJ* 2009;339:b2921.
19. World Health Organization Collaborative Study of Cardiovascular D, Steroid Hormone C. Venous thromboembolic disease and combined oral contraceptives: results of international multicentre case–control study. *Lancet* 1995;346:1575–82.
20. Stegeman BH, Bastos Md, Rosendaal FR, *et al*. Different combined oral contraceptives and the risk of venous thrombosis: systematic review and network meta-analysis. *BMJ* 2013;347: f5298.
21. NICE. Venous thromboembolic diseases: the management of venous thromboembolic diseases and the role of thrombophilia testing. NICE clinical guideline 144. In: NICE, ed.: NHS Evidence, 2012.
22. Hippisley-Cox J, Hammersley V, Pringle M, *et al*. How useful are general practice databases for research? Analysis of their accuracy and completeness in one research network. *Health Inform J* 2004;10:91–109.
23. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of non-gastrointestinal cancers: series of nested case-control studies using QResearch and CPRD data. *Br J Cancer* 2013;109:795–806.
24. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of gastrointestinal cancers: series of nested case-control studies with QResearch and CPRD data. *BMJ* 2013;346:f114.
25. Parker C, Coupland C, Hippisley-Cox J. Antipsychotic drugs and risk of venous thromboembolism: nested case-control study. *BMJ* 2010;341:c4245.
26. Sultan AA, West J, Tata LJ, *et al*. Risk of first venous thromboembolism in and around pregnancy: a population-based cohort study. *Br J Haematol* 2012;156:366–73.
27. Kuhnz W, Staks T, Jütting G. Pharmacokinetics of cyproterone acetate and ethinylestradiol in 15 women who received a combination oral contraceptive during three treatment cycles. *Contraception* 1993;48:557–75.
28. Dore DD, Norman H, Loughlin J, *et al*. Extended case–control study results on thromboembolic outcomes among transdermal contraceptive users. *Contraception* 2010;81:408–13.
29. Royal College of Obstetricians and Gynecologists. Venous Thromboembolism and Hormonal Contraception. Gree-top Guideline No. 40. 2010.
30. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: prospective cohort study. *BMJ* 2011;343:d4656.
31. Heit JA. The epidemiology of venous thromboembolism in the community: implications for prevention and management. *J Thromb Thrombolysis* 2006;21:23–9.
32. Pomp ER, Rosendaal FR, Doggen CJM. Smoking increases the risk of venous thrombosis and acts synergistically with oral contraceptive use. *Am J Hematol* 2008;83:97–102.
33. White RH, Zhou H, Romano PS. Incidence of idiopathic deep venous thrombosis and secondary thromboembolism among ethnic groups in California. *Ann Intern Med* 1998;128:737–40.
34. Bird ST, Hartzema AG, Brophy JM, *et al*. Risk of venous thromboembolism in women with polycystic ovary syndrome: a population-based matched cohort analysis. *Can Med Assoc J* 2013;185:E115–20.
35. The Health and Social Care Information Centre. *NHS contraceptive services: England, 2011/12. Community contraceptive clinics*. NHS HSCIC, 2012.
36. Royston P. Multiple imputation of missing values. *Stata J* 2004;4:227–41.
37. Office for National Statistics. *Opinions Survey Report No 41. Contraception and sexual health*. London: Office for National Statistics, 2009.
38. Herrett E, Thomas SL, Schoonen WM, *et al*. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010;69:4–14.
39. Lawrenson R, Todd JC, Leydon GM, *et al*. Validation of the diagnosis of venous thromboembolism in general practice database studies. *Br J Clin Pharmacol* 2000;49:591–6.
40. French R, Mercer C, Johnson A, *et al*. Use of contraceptive services in Britain: findings from the second National Survey of Sexual Attitudes and Lifestyles (Natsal-2). *J Fam Plann Reprod Health Care* 2009;35:9–14.

Paper 8

Vinogradova Y, Coupland C, Hippisley-Cox J. Combined oral contraceptives and risk of venous thromboembolism: nested case-control studies using the QResearch and the CPRD databases. *BMJ*. 2015;350:h2135.

BMJ 2015;350:h2135 doi: 10.1136/bmj.h2135

RESEARCH

Use of combined oral contraceptives and risk of venous thromboembolism: nested case-control studies using the QResearch and CPRD databases

Yana Vinogradova *research fellow in medical statistics*, Carol Coupland *associate professor and reader in medical statistics*, Julia Hippisley-Cox *professor of clinical epidemiology and general practice*

Division of Primary Care, University Park, Nottingham, NG2 7RD UK

Abstract

Objective To investigate the association between use of combined oral contraceptives and risk of venous thromboembolism, taking the type of progestogen into account.

Design Two nested case-control studies.

Setting General practices in the United Kingdom contributing to the Clinical Practice Research Datalink (CPRD; 618 practices) and QResearch primary care database (722 practices).

Participants Women aged 15-49 years with a first diagnosis of venous thromboembolism in 2001-13, each matched with up to five controls by age, practice, and calendar year.

Main outcome measures Odds ratios for incident venous thromboembolism and use of combined oral contraceptives in the previous year, adjusted for smoking status, alcohol consumption, ethnic group, body mass index, comorbidities, and other contraceptive drugs. Results were combined across the two datasets.

Results 5062 cases of venous thromboembolism from CPRD and 5500 from QResearch were analysed. Current exposure to any combined oral contraceptive was associated with an increased risk of venous thromboembolism (adjusted odds ratio 2.97, 95% confidence interval 2.78 to 3.17) compared with no exposure in the previous year. Corresponding risks associated with current exposure to desogestrel (4.28, 3.66 to 5.01), gestodene (3.64, 3.00 to 4.43), drospirenone (4.12, 3.43 to 4.96), and cyproterone (4.27, 3.57 to 5.11) were significantly higher than those for second generation contraceptives levonorgestrel (2.38, 2.18 to 2.59) and norethisterone (2.56, 2.15 to 3.06), and for norgestimate (2.53, 2.17 to 2.96). The number of extra cases of venous thromboembolism per year per 10 000 treated women was lowest for levonorgestrel (6, 95% confidence interval 5 to 7) and norgestimate (6, 5 to 8), and highest for desogestrel (14, 11 to 17) and cyproterone (14, 11 to 17).

Conclusions In these population based, case-control studies using two large primary care databases, risks of venous thromboembolism associated with combined oral contraceptives were, with the exception of norgestimate, higher for newer drug preparations than for second generation drugs.

Introduction

About 9% of women of reproductive age worldwide use oral contraceptives. This percentage rises to 18% of women in developed countries and 28% of women in the United Kingdom.[1] Combined oral contraceptives form a substantial proportion of these, particularly in more developed nations. Although combined oral contraceptives are generally effective in preventing pregnancy, they have measurable side effects such as venous thromboembolism (VTE). VTE is important, not only because of the prolonged time over which women might be exposed to such contraceptives, but also because VTEs are potentially avoidable and can be fatal.

Previous studies have shown varying risks for different types of oral contraceptives (such as third generation pills compared with first or second generation pills), but such studies were done some years ago,[2-6] and tended not to include new preparations containing drospirenone. Also, previous studies have generally had insufficient power to analyse the risks for more recent formulations[7-10] such as norgestimate. Few studies—only four of those referenced here[9 11-13]—have included any detailed analyses of dosage and, of these, only Lidegaard and colleagues[12] have covered a full range of prescribed drugs. Some studies did not control for all potential confounders (such as body mass index or smoking),[12] while others analysed only healthy users.[4 11 14] Different methodological approaches in studies have also made it difficult to compare and combine the results.[15]

Correspondence to: Y Vinogradova yana.vinogradova@nottingham.ac.uk

Extra material supplied by the author (see http://www.bmj.com/content/350/bmj.h2135?tab=related#datasupp)

Web appendix 1: Supplementary tables
Web appendix 2: READ codes of conditions affecting venous thromboembolism

Therefore, although the increased VTE risk associated with combined oral contraceptive drugs is established, the relative risks associated with different combinations remain inconclusive, especially for newer formulations.[16 17]

The UK has some of the largest sources of routinely collected data in the world, with longitudinal primary care records spanning up to 25 years and linked to secondary care data and mortality records. These databases cover many millions of patients, include data both on exposure and outcomes, and therefore are representative of the setting in which drugs are used. This makes the databases ideally suited to large scale safety studies of commonly used drugs.[18 19] In this study, we have used the two largest of these databases, QResearch (www.qresearch.org) and Clinical Practice Research Datalink (CPRD, www.cprd.com). Both have been used for earlier studies of associations between drug prescribing and VTE risks.[4 5 10 14 20 21]

Our objective was to quantify the associations between use of combined oral contraceptives and risk of VTE, adjusting for comorbidities and other available confounding factors. In particular, we were interested to analyse risks associated with newer or less used preparations such as drospirenone or norgestimate, quantify risks associated with various types of progestogen, and analyse the effect of different doses of oestrogen on VTE risks. To make the study more comparable with previous studies, we also replicated analyses for different subgroups by age and health status and for VTE cases with anticoagulation prescriptions.

Methods

Study design

The protocol for this study has already been published.[15] We undertook two similar studies using the CPRD (January 2014 version; 618 UK general practices) and QResearch database (version 38; 722 general practices) to quantify the association between prescribing of combined oral contraceptives and risk of incident VTE. We identified open cohorts of all women who had no records of VTE before the study, were aged 15-49 years, and were registered with the study practices between 2001 and 2013. Within each cohort, we designed two nested case-control studies with incident cases of VTE during the study period. This design was chosen as the most practicable, because it allowed us to work within the maximum extraction capabilities of the databases without losing any of the available cases—and therefore not compromising either the power of the study or the generalisability of the findings.[22]

The methods used in the study followed exactly those of the published protocol, with one difference related to the use of linked data. With respect to case identification, the protocol specified that "the main analysis will be run on all cases with VTE identified from the general practice data." QResearch is, however, closely linked at the individual patient level to hospital admissions data, and mortality records from the UK Office for National Statistics (ONS, www.ons.gov.uk/; complete for 99.8% of patients in QResearch, 99.9% of ONS mortality records, and 98% of hospital admissions records)[23 24]. So we identified VTE cases if, in QResearch, there was a relevant clinical code in the GP record, linked hospital record, or linked mortality record (web table 1), using the earliest recorded date on any of the three sources as the index date. For CPRD, however, not all practices were linked to these external data, so we could use only general practice records to identify VTE cases in CPRD.

For both databases, we matched each case to up to five controls by year of birth and from the same practice using incidence density sampling. Each control was allocated an index date, which was the date of first VTE diagnosis for the matched case. Eligible women had to have been registered with their practice for at least one year before the index date.

Because records of prescriptions for anticoagulant therapy (BNF 2.8.2) might indicate a previous VTE episode that was not recorded, cases with such records six or more weeks before the index date and controls with such records at any time before the index date were excluded from the analysis. We also excluded women if they had conditions such as oophorectomy, hysterectomy, and sterilisation, which normally preclude use of combined oral contraceptives. Women identified as pregnant or in the first three months after delivery at the index date were excluded, because they were less likely to be users of combined oral contraceptives and have an increased risk of VTE.[25] Cases or controls with conflicting prescriptions—two or more prescriptions for different combined oral contraceptives issued on the same date for the month before the index date—were also removed from the analysis.

Exposure to oral contraceptive drugs

Exposure to hormonal contraceptive drugs was based on prescription information in the last year before the index date. The main focus of the study was on individual combined oral contraceptives, which included all the most commonly used preparations in the UK: norethisterone, levonorgestrel, norgestimate, desogestrel, gestodene, and drospirenone (BNF 7.3.1). We included cyproterone, a hormonal treatment for acne, because it is also used as an oral contraceptive owing to its progestogen-like effect on the release of testosterone by the ovaries (BNF 13.6.2). For confounder control, the analysis included oral progestogen only contraceptives (BNF 7.3.2) and non-oral hormonal contraceptives (BNF 7.3.1 and BNF 7.3.2: implants, injections, transdermal patches, intrauterine and vaginal devices).

We investigated the recency of use by calculating the gap in days between the estimated date for the last use of a combined oral contraceptive and the index date, and categorising it as follows: used at index date or last use 1-28 days before the index date (current use); last use 29-365 days before the index date (past use); or no use in the last year before the index date. If a woman was exposed to more than one combined oral contraceptive in the last 28 days, only the latest time used was considered, but an indicator that she had switched type of oral contraceptive in the last 28 days was included in the analysis. No use in the last year was a reference category for all analyses unless otherwise stated.

We included the category of past use in the analysis to allow for women having an increased VTE risk associated with previous drug use, either because of a very recent cessation of exposure close to the start of the current use period or because of a delayed start of drug use from a previous prescription, such that some women classified as past users were actually current users. This approach was used only to approximate short term residual and misclassification effects, and should not be interpreted as a measure of long term residual risk. To emphasise this, we have reported odds ratios for past users only in the web tables.

Use of other hormonal contraceptives (oral progestogen only and non-oral hormonal treatments) was similarly categorised into current and past exposure and added to the analysis as confounders. We aggregated the data for combined and progestogen only non-oral contraceptives, because the numbers of current users for combined non-oral contraceptives were low

(13 cases and 24 controls in CPRD, 11 cases and 14 controls in QResearch) and lacked power for separate analysis.

Because VTE risk is likely to be highest in the first three months of oral contraceptive use,[26] we estimated the effect of duration of exposure on current users. We assessed exposure duration by calculating the number of days of exposure within the previous year. If the gap between the end of one prescription and the start of the next was 30 days or less, we considered exposure was continuous and combined the durations of the prescriptions. If a gap was longer than 30 days, only the latest period of exposure was considered.

Length of exposure duration was based on a period of 84 days, the most common length of a contraceptive prescription and also close to the end of the period of highest VTE risk associated with contraceptive use in other studies.[7 9] We classified duration as short term (≤84 days) and long term (>84 days), and combined it with recency of use into the following categories: short term current users (new users and restarters), long term current users (prevalent users), past use, and no use in the previous year.

In our samples, three contraceptives—norethisterone, desogestrel and gestodene—were prescribed in combinations having different doses of oestrogen. Owing to evidence of associations between higher VTE risks and higher doses of oestrogen[12], we undertook a further analysis of current users and categorised separately the oestrogen dose for these preparations (low dose (20 µg), normal dose (30-40 µg)), based on their most recent prescriptions before the index date. There was only one preparation with a high oestrogen dose (50 µg), which was combined with norethisterone. However, since there were only seven current users with this high dose preparation across both databases (one case and one control in CPRD, one case and four controls in QResearch), we included these women in the normal dose category. For all other drugs, only normal dose combinations had been prescribed.

Confounding factors

We identified the conditions affecting risk of VTE from the UK's health service guidelines related to VTE and hormonal contraceptives (web appendix 2).[27] Since these conditions might affect the prescribing decisions of doctors, we decided to adjust for these in all analyses. The chronic conditions for any patient had to be recorded before the index date in, to be included. These conditions were cancer, congestive cardiac failure, varicose veins, cardiovascular disease, rheumatoid arthritis, systemic lupus erythematosus, chronic renal disease, asthma, chronic obstructive pulmonary disease, Crohn's disease or ulcerative colitis, and coagulation disturbances (Leiden factor V, protein C and S deficiencies).

We also included traumatic events and events leading to immobilisation if recorded in the six months before the index date. These events included acute infections (upper and lower respiratory tract infections, urinary tract infections), surgery or leg/hip fracture, admission to hospital (excluding the previous 30 days before the index date). Non-idiopathic groups were formed from women with any of these chronic conditions or events, and idiopathic groups from women without them.

Obesity and smoking are also mentioned as potential risk factors in the NHS guidelines, so we adjusted all analyses for body mass index as a continuous variable, and for smoking status as the following categories: current smoker (light (1-9 cigarettes/day), medium (10-19), heavy (≥20); ex-smoker; non-smoker. We used values recorded at the closest date before the index date.

We included polycystic ovary syndrome as a confounder because it is treated with hormonal contraceptives and associated with an increased risk of VTE.[28] Other conditions treated with hormonal contraceptive prescriptions—acne, hirsutism, and menstrual disorders—were initially considered as potential confounders but their addition to analyses failed to change odds ratios for main exposures by more than 10%, so these were not included in the final study analyses.

Alcohol consumption has previously been considered as a confounder[10] and, being a potentially important lifestyle factor available from primary care data,[29] was categorised and included in the analyses (light (≤2 units/day), medium to heavy (≥3), ex-use or no use). We also adjusted for ethnic group (white or not recorded, Asian, black, or other), because women in ethnic minorities could have different patterns of contraceptive use[30] and different VTE risks from the white population.[31]

Social deprivation, which can be measured in the UK by the Townsend score, was not included as a confounder in the main analyses because it was not a significant risk factor for VTE in a previous QResearch study.[32] Furthermore, the CPRD had a large proportion of missing data for the Townsend score, so the inclusion of social deprivation would result in a loss of statistical power in that analysis. However, during the peer review process, we decided to run an additional analysis on QResearch data including the Townsend score as a confounder, because the Townsend data were almost complete (available for 99.8% of cases and controls). We have, therefore, run an additional analysis on QResearch data including the Townsend score as a confounder.

Statistical analysis

The analyses were run on each database separately. Crude incidence was calculated by dividing the number of cases with incident VTE by the number of person years in the cohorts. Data for oral contraceptive exposure were only available for cases and matched controls rather than whole cohorts, which had higher proportions of older women than the general population. Therefore, we estimated age standardised rates of exposure to any oral contraceptives, using groups of controls before exclusions and directly standardising to the age profile for the UK general population in the relevant year based on data from the UK Office for National Statistics.

We used conditional logistic regression to obtain odds ratios with 95% confidence intervals. The differences between exposures were assessed using Wald's tests. To account for the log normal distribution for body mass index, we used the logarithm of body mass index for all analyses. Missing values for body mass index, smoking status, and alcohol consumption were imputed using chained equations.[33] Ten imputed sets were generated, and the imputation model included age, outcome (case or control), index year, all confounding factors (including acne, hirsutism, and menstrual disorders), exposure to progestogen only oral contraceptives, non-oral contraceptives (progestogen only and combined), and recency and duration of use for combined oral contraceptives. We combined the results from the imputed sets using Rubin's rules.[33]

To facilitate comparison of our results with those from earlier studies, which had analysed the associations of exposure to combined oral contraceptives by reference to levonorgestrel, we reran the analyses comparing current exposure to each drug of interest with current exposure to levonorgestrel (in combination with a normal oestrogen dose (30-40µg), the only doses prescribed in our data). Current exposures to levonorgestrel and the drug of interest were replaced with a

variable coded as exposure to the drug, no exposure to the drug, and exposure to levonorgestrel. Analyses were adjusted for past exposure to levonorgestrel and the drug of interest, exposure to other combined oral contraceptives, and confounding factors.

We ran three additional analyses to look at methodological issues and allow comparisons with other published studies. Because results of diagnostic tests for VTE are not generally included in the primary care electronic records, some studies[11 14] used subsequent anticoagulation therapy to confirm VTE diagnosis, including only patients treated as such despite possible under ascertainment of VTE cases. In our study, anticoagulation records were available only for prescriptions in primary care, representing doctors' initial responses to patients presenting with VTE symptoms rather than a more complete record of initial and subsequent treatments. However, to facilitate comparison with these studies, we ran another analysis on VTE cases, supported with either prescriptions for anticoagulation therapy (BNF 2.8.2) or records of death within six weeks of the recorded date of VTE diagnosis. Links to individual mortality data from the ONS were available for all QResearch practices, so these were included in identification of deaths due to VTE. This was not the case for CPRD practices, however, so identification of deaths for the CPRD analysis was derived solely from the general practitioner record.

To distinguish whether there are different associations in idiopathic cases compared with non-idiopathic cases, an additional stratified analysis was run on subgroups of cases and matched controls. In this analysis, idiopathic cases were first analysed with any idiopathic matched controls (that is, controls with none of the chronic conditions or events listed above). Then, only non-idiopathic cases were analysed with any non-idiopathic matched controls (that is, controls with one or more of the chronic conditions or events used to identify non-idiopathic cases). The third analysis was run on subgroups of younger (15-24 years) and older (25-49 years) women, because younger women are more likely to use contraceptive clinics as a source of oral contraceptives, potentially leading to a lack of recorded exposure data for this group.[30]

In the protocol, we had proposed a sensitivity analysis for practices linked to hospital admission data, where VTE cases would be identified not only from the practice records but also from hospital admissions data. For QResearch, because the selection process used linked data sources including hospital admissions, this additional analysis became redundant. Instead, we ran a sensitivity analysis using QResearch cases identified only through general practice medical records and matched controls. For CPRD, we ran the proposed sensitivity analysis for data from the subset of practices linked to both hospital admission data and ONS mortality data, where data from all sources were used to identify VTE cases. VTE cases in hospital admission and ONS mortality data were identified by ICD-10 codes (web table 1).

To increase the power of the study and obtain more precise estimates, we combined results from the two databases using a meta-analysis technique. Adjusted odds ratios from the conditional logistical regression analyses of the two datasets were pooled by use of a fixed effect model with inverse variance weights.[34] We chose a fixed effect model because—apart from the necessarily different approaches to identification of relevant cases described above—the studies in CPRD and QResearch (which have similar sizes and similar methods of recording information) were comparable, using the same exclusion criteria, definitions of exposures and confounders, and the same models. In view of these similarities, differences in observed associations seemed most likely to derive from sampling variations, but we also ran a sensitivity analysis using a random effect model to allow for any heterogeneity.

To estimate the magnitude of VTE risk associated with combined oral contraceptives, we calculated the numbers needed to harm per year by using the adjusted odds ratios derived from the combined analyses.[35] The incidence for the unexposed female population could not be derived either from QResearch or CPRD because exposure details were not available for the whole cohorts. The rate was, therefore, derived from a Danish cohort[12] taking into account the differences in study design. We based our calculations for numbers needed to harm on the adjusted odds ratios from the combined analyses for current use and the Danish study rates of 4.18 per 10 000 women years for women aged 15-49 years and 4.91 per 10 000 women years for those aged 25-49 years. We also estimated the number of additional VTE cases expected per year per 10 000 treated women.

We used Stata version 13 for the analyses. All available cases were used from both QResearch and CPRD. A 1% level of statistical significance was used to account for multiple comparisons and 95% confidence intervals to enhance comparability with other studies. For clarity, only odds ratios from the combined analyses are presented and discussed, but the contributing odds ratios from CPRD and QResearch can be found in the tables.

Results

We identified 7334 incident VTE cases from CPRD based on clinical Read codes recorded in the general practitioner data, and 8211 incident VTE cases from QResearch within the study period, both with at least one year of medical records. Crude incidence of VTE cases per 10 000 women years was 5.9 (95% confidence interval 5.7 to 6.0) in CPRD and 6.1 (6.0 to 6.3) in QResearch. After matching cases to controls and removing ineligible participants, the final analysis included 5062 (69%) VTE cases from CPRD matched to 19 638 controls, and 5500 (67%) VTE cases from QResearch matched to 22 396 controls (fig 1⇓). Of 5500 VTE cases from QResearch, 5088 (93%) were identified from primary care records, and an additional 284 (5%) from hospital admission data and 128 (2%) from ONS mortality data. For CPRD cases, 2917 (58%) VTE events were recorded as deep vein thrombosis only; 1626 (32%) as pulmonary embolism, with or without deep vein thrombosis; and 519 (10.3%) as other types of VTE; corresponding numbers for QResearch cases were 3156 (57%), 1613 (29%), and 731 (13%).

Proportions of cases and controls across the demographic measures and morbidities relevant to the study showed the similarities between database populations (table 1⇓, web table 2). Median ages of women in the study were 38 years (interquartile range 30-44) for CPRD and 39 years (31-44) for QResearch. Current smoking was more common in cases than controls (27% *v* 21% for both databases), as was obesity (body mass index≥30; 30% *v* 17% for CPRD, 24% *v* 14% for QResearch). Proportions of women with established risk factors for VTE (that is, non-idiopathic cases and controls) were similar for each database (47% cases and 27% controls for CPRD, 47% and 26% for QResearch). About half of women with VTE in the study had anticoagulation prescriptions or died within six weeks of the recorded diagnosis date (2454 and 79 cases, respectively, or 50% overall in CPRD; 2749 and 207, or 54% overall in QResearch).

Exposure, main analysis

Age standardised rates of exposure to any oral contraceptive did not change over the study period (overall rates 29% in

CPRD, 26% in QResearch). Use of levonorgestrel, the most common combined oral contraceptive, decreased during the study (from 15% to 11% in CPRD, and 13% to 10% in QResearch), whereas use of progestogen only oral contraceptives rose from 3% to 7% (fig 2⇓).

In the year before the index date, 30% of cases and 18% of controls in CPRD had at least one prescription for combined oral contraceptives. For QResearch, the numbers were 28% of cases and 16% of controls. Preparations with levonorgestrel seemed to be the most commonly prescribed combined oral contraceptives (45% of exposed cases, 54% of exposed controls in CPRD; 44%, 52% in QResearch). Other contraceptive types were much less used (all between 7% and 13%). Most users of combined oral contraceptives within the previous year were current users—that is, exposed in the last 28 days (84% of exposed cases, 79% of exposed controls in CPRD; 84%, 77% in QResearch; web table 3). Most of the current users were exposed for more than 84 days (across different permutations of drug type, database, and cases and controls, all between 70% and 87%).

For the analyses combining CPRD and QResearch results, current use of any combined oral contraceptive was associated with a significantly increased VTE risk (adjusted odds ratio 2.97, 95% confidence interval 2.78 to 3.17) compared with no exposure in the last year. The risks varied between different types of oral contraceptives and resulted in two clear groups: norethisterone, levonorgestrel, and norgestimate in one group; and desogestrel, gestodene, drospirenone, and cyproterone in the other. Current exposure showed that the first group had a two and a half times increased VTE risk (levonorgestrel (2.38, 2.18 to 2.59), norethisterone (2.56, 2.15 to 3.06), and norgestimate (2.53, 2.17 to 2.96), and roughly a four times increased risk for the second group (desogestrel (4.28, 3.66 to 5.01), gestodene (3.64, 3.00 to 4.43), drospirenone (4.12, 3.43 to 4.96), and cyproterone (4.27, 3.57 to 5.11) all compared to no exposure in the last year (table 2⇓, fig 3⇓, web table 4 for all variables in the model).

In our analysis to facilitate comparison with existing studies, risks associated with current use of norethisterone and norgestimate did not differ significantly from levonorgestrel. However, the risk associated with current use of gestodene was 1.5 times higher than for levonorgestrel (adjusted odds ratio 1.52, 95% confidence interval 1.24 to 1.87) and about 1.8 times higher for desogestrel, drospirenone, and cyproterone (table 3⇓).

Analyses of oestrogen dosages were possible only for norethisterone, desogestrel, and gestodene (20 µg; 30-40 µg). Desogestrel was the most commonly prescribed of these three drugs and had slightly higher odds ratios for higher doses, whereas norethisterone and gestodene had higher odds ratios for lower doses; however, none of these differences between doses was significant (table 2).

Analysis of the duration for current users showed, only for levonorgestrel, a significantly increased risk for new users and restarters (that is, short term users) compared with long term users (adjusted odds ratios 3.38 (95% confidence interval 2.86 to 3.99) *v* 2.16 (1.97 to 2.38), P<0.001). For other drug types, the results were inconsistent, with odds ratios for shorter exposure marginally higher for norethisterone and gestodene, but marginally lower for norgestimate, desogestrel, drospirenone, and cyproterone (web table 5). Adjusted odds ratios for other confounders, including use of other hormonal contraceptives (oral progestogen only and non-oral hormonal treatments) and associations for our category of past use, are available in web table 4.

Although previous studies have not shown any confounding effect from body mass index,[11] we found that inclusion of body mass index into the model changed odds ratios for drug exposures by percentages ranging from 7% to over 10%, with the highest effect for drospirenone (web table 6). Each risk factor, when included individually, did not show a major effect on the results for drug exposures. But when all combined, the odds ratios for drug exposures changed by percentages of between 13% and 25% compared with the unadjusted values. Adjustment for deprivation information in QResearch changed odds ratios for exposures by up to 5%.

Additional analyses

When restricted to cases with anticoagulation prescriptions and matched controls, the overall pattern of risks was similar to those from the main analysis (table 4⇓), although odds ratios were higher for all combined oral contraceptive drug types within a wide range of relative change. The differences were smaller for norethisterone (8% increase in adjusted odds ratio) and levonorgestrel (24%), and larger for norgestimate (40%), gestodene (78%), desogestrel (46%), drospirenone (48%), and cyproterone (40%). However, when tabulated by exposure, the variations shown in proportions of cases with anticoagulation prescriptions for different exposure groups (web table 7) might reflect some differential treatment of patients at initial presentation based on known drug risks.

The analysis for idiopathic cases (that is, with no risk associated conditions) and matched controls showed higher odds ratios for the oral contraceptives in the idiopathic analysis than the main analysis (table 4), but odds ratios by type of oral contraceptive were similar to the main analysis results. The odds ratios for the non-idiopathic group were correspondingly smaller (web table 8), but not as reliable because fewer non-idiopathic controls were available to match to non-idiopathic cases, leading to a reduction of the original matching ratio of cases to controls from 1:5 to about 1:1.5.

In the analysis of VTE cases according to age group, the proportion of the younger group was small (15-24 years; 13% in CPRD, 9% in QResearch). Odds ratios were lower for this group than for the older group (25-49 years; table 4), but again the overall pattern of risk stayed in line with the main analysis.

Risks for combined oral contraceptives compared with levonorgestrel were consistent across all the additional analyses (table 5⇓), with no significant differences for norethisterone and norgestimate. Odds ratios for other drugs ranged from 1.4 to 2.4 (all significant apart from some drugs in the non-idiopathic group and in the younger group, which were likely to be due to low numbers).

The results from CPRD and QResearch were similar with the exception of those for norgestimate. In the CPRD analyses, risks associated with norgestimate use were similar to risks for levonorgestrel, whereas in the QResearch analyses, risks for norgestimate consistently fell between those for levonorgestrel and desogestrel across all analyses. However, the combined results, which gave more precise estimates, placed norgestimate in the group with levonorgestrel and norethisterone. An additional analysis for QResearch, which included adjustment for the Townsend deprivation score, showed results similar to the main analysis (web table 9).

Sensitivity analyses

When combining the results from the databases we discovered significant heterogeneity only for current use of norgestimate (I^2=89%, P=0.003). The direction of the effect was the same in both databases and, after we applied a random effect model to combine the results, the estimate for norgestimate did not change our conclusion of its association being close to the estimates for the group of earlier contraceptives (combined odds ratio 2.49, 95% confidence interval 1.56 to 3.97).

The sensitivity analysis for QResearch cases identified only through general practice medical records and matched controls delivered results in line with the main analysis (web table 10). The sensitivity analysis for CPRD practices linked to hospital admission and ONS mortality data was based on 346 general practices and covered the period between 1 January 2001 and 30 March 2012. The crude incidence of VTE per 10 000 women years in this cohort was 5.7 (95% confidence interval 5.5 to 5.8). We identified 436 extra cases from hospital admission data and 14 from ONS mortality data with at least one year of medical records. After exclusions, 2989 cases were included in the analysis, of which 2654 (89%) were identified from general practice records, 324 (11%) from hospital admission data, and 11 (0.4%) from ONS mortality data. The results were also in line with the main analysis (web table 11).

Numbers needed to harm and excess risk

Because combined oral contraceptive use was associated with increased VTE risk, additional cases of VTE would be expected across all types of combined oral contraceptives in exposed women compared with unexposed women, and particularly in those aged 25-49 years (table 6⇓). The lowest numbers of extra cases of VTE per year per 10 000 treated women were six extra cases for levonorgestrel (6, 95% confidence interval 5 to 7) and norgestimate (6, 5 to 8) for women aged 15-49 years, and seven extra cases for levonorgestrel (7, 6 to 8) and norgestimate (7, 5 to 9) for those aged 25-49 years. The highest numbers of extra cases of VTE per year per 10 000 treated women were for desogestrel (14 extra cases, 11 to 17) and cyproterone (14, 11 to 17) for ages 15-49 years, and for drospirenone (17, 13 to 23), desogestrel (17, 13 to 21), and cyproterone (17, 12 to 22) for ages 25-49 years.

Discussion

In this observational study based on two large primary care databases, women exposed to drospirenone, gestodene, cyproterone, and desogestrel within the last 28 days had around a four times increased risk of VTE. Women exposed to levonorgestrel, norethisterone, and norgestimate had about two and a half times increase in VTE risk compared with women not exposed in the past year. Risks for current use of gestodene, drospirenone, cyproterone, and desogestrel were 1.5-1.8 times higher than for levonorgestrel. Results from the additional analyses stayed in line with the main findings, although there were stronger associations in the analyses restricted to cases with anticoagulant prescriptions and matched controls. These differences were expected and can be explained by our methodological approach. We saw no significant association in the analyses of oestrogen dosages.

Strengths and limitations of the study

The main strengths of this study are its recency, comprehensiveness, and generalisability. It was based on the general female population in the UK aged 15-49 years, and explored exposure to combined oral contraceptives commonly prescribed during the past 13 years. The study also benefitted from the statistical power of large samples from the two largest UK primary care databases. Consistency in records for diagnoses, lifestyle information, and prescriptions allowed us to combine the results from both databases and achieve narrower confidence intervals for our estimates. The study also benefitted from a consistent design.

Results were adjusted for several confounding factors such as body mass index, smoking status, alcohol use, and social deprivation, which were not available to some previous studies. Education and family history might also be considered to be confounders but neither could be included in the analysis because they are not recorded sufficiently often on either the QResearch or CPRD databases. Because the exposure was based on systematically recorded prescription information, the study was free from recall bias. All eligible women were included, thus eliminating selection bias. Several additional analyses looking at conflicting methodological issues from previous studies allow readers to compare and assess the validity of the results.

A study limitation was the potential misclassification of exposure to combined oral contraceptives. According to the Contraception and Sexual Health survey in Great Britain (2000-09), between 25% and 28% of women used an oral contraceptive depending on the year.[36] Our data for both databases had similar age standardised rates of exposure to any oral contraceptive—26% for QResearch and 29% for CPRD. Because exposure information is based on prescriptions, however, there is a degree of uncertainty about actual use—when a woman started taking the drug or whether she took it at all. According to one survey from the United States, 19% of women discontinued using oral contraceptives within the first six months, more commonly younger women.[37] Because outcome information was collected prospectively, however, we do not see any reason why this effect should differ between cases and controls. Such misclassification of exposure might, however, shift odds ratios towards unity. Some uncertainty also relates to women who may have delayed use of drugs from past prescriptions (and so were actually current rather than past users), and to unaccounted residual risk associated with women who ceased use for any reason just before the current use period. However, these two potential misclassifications are likely to be small.

NHS community contraceptive clinics are also a source of oral contraceptive pills apart from general practice doctors. According to NHS Contraceptive Services reports issued between 2005 and 2013 (www.hscic.gov.uk), on average 6.9% of women under 25 years old and 1.6% of older women received oral contraceptive pills from contraceptive clinics. One report in 2005 released the numbers separately for combined and progestogen only pills, showing that the proportion of combined contraceptives prescribed was 91% of all oral contraceptives for younger women and 73% for older women.[38] From these figures, we estimated that in the population, 6.3% of younger and 1.2% of older women had exposure to combined oral contraceptives without related general practice records. These women would appear in our analyses as not exposed, creating a potential underestimation that might shift odds ratios towards unity, with an effect likely to be greater in the younger group.

The additional analyses for younger women did, in fact, produce lower odds ratios for all drugs apart from levonorgestrel and norgestimate. However, in the direct comparisons of different oral contraceptives with levonorgestrel, there was no potential bias with respect to misclassification of non-users because only

oral contraceptive users were involved. Other biases could arise if the prescribing regimens of contraceptive clinics differed markedly from those of general practices (with one or other being more inclined towards higher risk, lower priced drugs), or if the material circumstances of women attending general practices differed from those attending contraceptive clinics. No published data seem to support this, however, and we believe that any such effects are likely to be negligible especially given the much higher proportion of supply from general practices.

There is also some degree of uncertainty in VTE diagnoses in both CPRD and QResearch practice records, because the results of diagnostic tests needed to confirm VTE are not generally available on the primary care databases. Furthermore, these diagnoses cannot be adjudicated in our study as might happen in a clinical trial, so may be subject to misclassification bias, with some false positives for cases and some false negatives for controls. The likelihood of misclassifications is, however, much higher for cases than controls because of the low incidence of VTE in the general population from which the controls are selected—therefore, overall, such errors and misclassifications if non-differential would tend to shift odds ratios towards unity.

Further, the incidence of VTE in our cohorts were both within the estimated range of five to 10 cases per 10 000 person years for young women.[39] The slightly higher rate within the QResearch cohort can be explained partly because the data used in the database's analysis was augmented by linked mortality information from the ONS and hospital episode statistics. This link will have added extra cases to the QResearch analysis and reduced diagnostic errors. However, the relatively small difference in rates between QResearch and CPRD, and the fact that the difference is also partly explained by the slightly higher median age of the QResearch cohort, suggests that neither analysis has been substantially affected by diagnostic errors. An earlier study has also shown that the addition of "possible" cases of VTE did not materially affect results obtained using only verified cases.[40]

Patients with a diagnosis of VTE are usually treated with anticoagulant medication. In our data, however, there are several reasons why VTE cases might not be followed by an anticoagulation prescription, such as a VTE event resulting in death, or treatment unrecorded in the GP record because it was initiated and continued in a hospital or other community setting. We found that, overall, about half of patients with VTE had a record of anticoagulation prescription within their general practice record. But a more detailed breakdown by exposure and drug type revealed possible differential treatment of exposed patients depending on contraceptive drug type and roughly reflecting the known VTE risks of the drugs.

The higher odds ratios in the additional analysis restricted to cases with anticoagulation prescriptions than those from the main analysis can be explained by a combination of the exclusion of uncertain events and differential anticoagulant prescribing by doctors. Women who receive anticoagulation treatment, which is necessary for VTE, are normally more likely to be true cases than those with no treatment recorded. Therefore, inclusion of some non-cases in our main analysis probably shifted odds ratios towards unity. On the other hand, our conjecture—based on evidence in our data of differential prescribing—is that doctors are more likely immediately to prescribe and record anticoagulants for patients with VTE symptoms exposed to a high risk oral contraceptive drug than for users of lower risk drugs. As a result, use of anticoagulation records to exclude uncertain events is more problematic in this study, and we would argue that results of our restricted analysis should be read with caution, indicating little more than a general agreement with earlier findings of increased odds ratios. In particular, the range of relative increases is probably exaggerated and comparisons between drug types possibly less reliable.

Finally, the higher odds ratios obtained from the subgroups with idiopathic cases and matched idiopathic controls, compared with odds ratios from the main analysis, were also expected because the absolute risk of VTE for unexposed patients is smaller in an idiopathic subgroup than that in a non-idiopathic subgroup (and by extension a general population).[41] Although the associations seem to be stronger in the idiopathic analysis, we do not believe that they are necessarily generalisable because of the wide variation in definitions of idiopathic groups across existing studies, and the general difficulties that have been noted in defining such groups.[42]

Comparison with recent studies

In our study, we observed a reduction in prescription rates for combined oral contraceptives and an increased rate for progestogen only oral contraceptives. This is in line with NHS statistics for prescriptions in the community, and might reflect the effects of various guidelines and recommendations for patients at high risk of VTE.[43]

Prior to our study, the largest study of VTE and combined oral contraceptives was a cohort study based on medical records from the Danish general population, covering the period 2001-09 and identifying 4246 women with a first recorded VTE.[12] The Danish study adjusted for age, calendar year, and level of education. By comparison, our study had more than twice the number of VTE cases; added a further four years of data; adjusted for body mass index, smoking status, alcohol consumption, ethnic group, several chronic and acute conditions associated with increased VTE risk, and use of other hormonal contraceptives; and accounted for age, calendar year, and practice by matching. Not all types of combined contraceptives in the Danish study were available for comparison, because some are rarely prescribed in the UK. The most used contraceptives were levonorgestrel in the UK and gestodene in Denmark. In our main analysis, the odds ratios for current use of available contraceptives were similar to the Danish relative rates. Despite a difference in the proportion of cases with anticoagulant prescriptions (52% in our study *v* 67% in the Danish study), results in these subgroups were also similar.

The most recent CPRD based study focused on a comparison of VTE risk in idiopathic cases of VTE with anticoagulant prescriptions between levonorgestrel and drospirenone.[14] It was run on records from 2002 to 2009, and so was based on fewer practices than in our study. For current users, that study showed a threefold increase in VTE risk for drospirenone compared with levonorgestrel (17 *v* 44 exposed cases; odds ratio 3.3, 95% confidence interval 1.4 to 7.6). In our study, the odds ratios for current use of drospirenone were about twice as high as for levonorgestrel in our main analysis and all additional analyses. Another study (2002-08),[11] based on pharmacological records from a US company and with a design similar to the recent CPRD study,[14] had more women with VTE exposed to drospirenone than levonorgestrel (121 *v* 65). It also showed an increased risk of VTE with drospirenone compared with levonorgestrel (odds ratio 2.4, 95% confidence interval 1.7 to 3.4). Based on the same source of data (the same US company),[11] another study showed a 70% increased risk associated with desogestrel (1.7 (1.1 to 2.4)) and no significant increase with norgestimate, both compared to levonorgestrel.[44] All three of these studies differed from ours in terms of case inclusion

criteria, but their results align well with those from our additional analyses.

An Austrian case-control study (2002-06)[45] investigated gestodene-containing and second generation oral contraceptives (79 and 83 exposed cases, respectively), identifying cases from referral centres and hospitals and deriving exposure information from questionnaires. Odds ratios for contraceptive use (with reference to non-users) were two to three times higher than in our study. But, as the authors suggested, this increased risk might be due to what they termed as "hospital bias," which can lead to overestimation of VTE risks.[46] The study also compared gestodene with second generation pills but did not show any significant difference between the drugs in several sensitivity analyses. The relative differences between levonorgestrel and gestodene seen in our main and additional analysis for idiopathic cases were within the confidence interval or close to the upper confidence levels of this study.

A Dutch study (1999-2004)[9] analysed all available oral contraceptives, identifying women with VTE from anticoagulation clinics and assessing exposure from postal questionnaires and interviews. Most controls were, however, acquired by random digit dialling, a technique that might have led to selective recruitment of a less active group with a poorer health profile than the general population.[47] This technique and the higher response rates in women with VTE than in those controls (79% *v* 64%) might have introduced a selection bias and inflated odds ratios. In fact, the study did report higher odds ratios that those more generally reported elsewhere and consistently higher odds ratios with reference to non-use than our study, although relative differences with reference to levonorgestrel were again close to our findings.

An Israeli cohort study[48] (2002-08) compared VTE risks for drospirenone with those for second and third generation oral contraceptives and found significant differences for drospirenone compared with both generations (rate ratio 1.65 (95% confidence interval 1.02 to 2.65), 1.43 (1.15 to 1.78), respectively). The pattern of prescribing in this study was different from ours, with most common exposure to third generation drugs (384 exposed cases) and a lower use of levonorgestrel (23 exposed cases). Our study showed a similar association for current use of drospirenone compared with levonorgestrel (odds ratio of 1.75), but found no difference between drospirenone and third generation drugs.

Despite being a third generation drug, norgestimate (282 exposed cases) had associations with VTE risk similar to levonorgestrel in our study. But because norgestimate partly metabolises to levonorgestrel,[49] its classification as a third generation drug is not clearly established. A Danish review classified norgestimate as a second generation drug and recommended prescribing it as a first choice contraceptive along with levonorgestrel and norethisterone.[50] Norgestimate has a lower androgenic effect than levonorgestrel and had been used at a similar level to levonorgestrel in the Denmark study,[12] although in our study levonorgestrel was prescribed three times more often than norgestimate. No significant difference between norgestimate and levonorgestrel was shown in the Danish study[12] (165 exposed cases, rate ratio 1.18 (95% confidence interval 0.86 to 1.62)) or in the US study[44] (124, odds ratio 1.1 (95% confidence interval 0.8 to 1.5)).

A meta-analysis[16] including the Danish and US studies also demonstrated this non-difference between norgestimate and levonorgestrel, although it was not highlighted in the main study findings, which focused on different drug generations and oestrogen dosages. Although norgestimate had been on the market from 1995, other studies either did not consider norgestimate or were underpowered (for norgestimate, only five exposed cases in the Dutch study,[9] 15 in the CPRD study,[4] and an unclear number in a German study with lower total numbers[13]).

Our study showed no associations between VTE risk and oestrogen dose for the three types of combined contraceptives, where this could be assessed. Levonorgestrel in the UK was prescribed mostly with a 30-40 µg dose of oestrogen, so oestrogen dose analysis was not possible. Comparable preparations for norethisterone have not been analysed before, so direct comparison of our results with other studies is not possible. A lower dose of oestrogen for desogestrel preparations was associated with a slightly lower risk of VTE, which was consistent with existing literature,[12 16] but our difference was not significant. For combinations with gestodene, the numbers of current users were insufficient to draw any meaningful conclusions.

Conclusion

This study, based on two large primary care databases, investigated risks of VTE associated with combined oral contraceptives prescribed to the general female population in the UK. We believe this study has the statistical power and sufficient adjustment for relevant confounders to be regarded as an important clarifying study, which has produced the most reliable possible risk estimates using currently available UK prescription data. It has confirmed results from other recent large scale studies and added new evidence, particularly for newer or less used combined oral preparations, such as those containing drospirenone or norgestimate. Risks associated with combined oral contraceptives were, apart from norgestimate, higher for newer drug preparations than for second generation drugs.

The results from our study and the Danish study[12] provide evidence for relevant authorities concerned with prescribing guidelines or those involved with regulation of safety of medicines. In particular, along with the Danish study and a US study,[44] our results confirm the similarity of risks for levonorgestrel and norgestimate in a UK context.

We acknowledge the contribution of Egton Medical Information System (EMIS) and the University of Nottingham for expertise in creating and maintaining QResearch and to the EMIS practices which contribute data; we thank CPRD and Vision Practices for allowing access to the CPRD for this study.

Contributors: JHC had the original idea for this study. CC contributed to the development of the idea and the study design. YV reviewed the literature, contributed to the study design, undertook the primary analysis as well as the first interpretation and wrote the first draft of the paper. JHC and CC critically reviewed the paper. All authors approved the submitted version.

Funding: This research received no external funding.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any additional organisation for the submitted work; JHC is professor of clinical epidemiology at the University of Nottingham and unpaid director of QResearch, a not-for-profit organisation which is a joint partnership between the University of Nottingham and EMIS (commercial IT supplier for 60% of general practices in the UK); JHC is also a paid director of ClinRisk, which produces open and closed source software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help improve

What is already known on this topic

Oral contraceptive pills are known to be associated with an increased risk of thromboembolism (VTE)

Despite comparing third generation contraceptive pills with first and second generation pills, previous studies have had insufficient power to quantify VTE risk with individual drugs, particularly for new or less commonly used preparations such as drospirenone or norgestimate

What this study adds

This study, based on national population and prescribing practices in the UK, has sufficient power to provide reliable comparative findings for different formulations of combined oral contraceptives; its findings are comparable to those based on a Danish national cohort study

Preparations containing gestodene, desogestrel, drospirenone, and cyproterone were associated with significantly higher risks of VTE than preparations containing either levonorgestrel or norgestimate

The number of extra VTE cases per year per 10 000 treated women was lowest for levonorgestrel and norgestimate, and highest for desogestrel and cyproterone

patient care; no other relationships or activities that could appear to have influenced the submitted work.

Ethics and dissemination: The protocol for this study has been published in BMJ Open. It has also been independently peer reviewed by the QResearch Scientific Board and has been reported to Trent research ethics committee in accordance with the agreed procedure (reference no MREC/03/4/021). For CPRD data analysis, the protocol was approved by Independent Scientific Advisory Committee (reference no ISAC 13_118RA2).

Data sharing: Results for all additional analyses and descriptive statistics are already published in the web tables. Any further requests are available from the corresponding author.

The lead author and the manuscript's guarantor (YV) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned and registered have been explained.

1 Department of Economic and Social Affairs. World contraceptive patterns 2013. United Nations, 2013.
2 Lewis M, Heinemann L, MacRae K, Bruppacher R, Spitzer W. The increased risk of venous thromboembolism and the use of third generation progestagens: role of bias in observational research. *Contraception* 1996;54:5-13.
3 Farmer RDT, Lawrenson RA, Thompson CR, Kennedy JG, Hambleton IR. Population-based study of risk of venous thromboembolism associated with various oral contraceptives. *Lancet* 1997;349:83-8.
4 Farmer RDT, Lawrenson RA, Todd JC, et al. A comparison of the risks of venous thromboembolic disease in association with different combined oral contraceptives. *Br J Clin Pharmacol* 2000;49:580-90.
5 Lawrenson R, Farmer R. Venous thromboembolism and combined oral contraceptives: does the type of progestogen make a difference? *Contraception* 2000;62(2 suppl):S21-8.
6 Heinemann LAJ, Lewis MA, Thorogood M, Spitzer WO, Guggenmoos-Holzmann I, Bruppacher R. Case-control study of oral contraceptives and risk of thromboembolic stroke: results from International Study on Oral Contraceptives and Health of Young Women. *BMJ* 1997;315:1502-4.
7 Dinger JC, Heinemann LAJ, Kühl-Habich D. The safety of a drospirenone-containing oral contraceptive: final results from the European Active Surveillance study on Oral Contraceptives based on 142,475 women-years of observation. *Contraception* 2007;75:344-54.
8 Seeger JD, Loughlin J, Eng PM, Clifford CR, Cutone J, Walker AM. Risk of thromboembolism in women taking ethinylestradiol/drospirenone and other oral contraceptives. *Obstet Gynecol* 2007;110:587-93.
9 van Hylckama Vlieg A, Helmerhorst FM, Vandenbroucke JP, Doggen CJM, Rosendaal FR. The venous thrombotic risk of oral contraceptives, effects of oestrogen dose and progestogen type: results of the MEGA case-control study. *BMJ* 2009;339:b2921.
10 Seaman HE, de Vries CS, Farmer RDT. Venous thromboembolism associated with cyproterone acetate in combination with ethinyloestradiol (Dianette): observational studies using the UK General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2004;13:427-36.
11 Jick SS, Hernandez RK. Risk of non-fatal venous thromboembolism in women using oral contraceptives containing drospirenone compared with women using oral contraceptives containing levonorgestrel: case-control study using United States claims data. *BMJ* 2011;342:d2151.
12 Lidegaard Ø, Nielsen LH, Skovlund CW, Skjeldestad FE, Løkkegaard E. Risk of venous thromboembolism from use of oral contraceptives containing different progestogens and oestrogen doses: Danish cohort study, 2001-9. *BMJ* 2011;343:d6423.
13 Lewis MA, MacRae KD, Kühl-Habich1 D, Bruppacher R, Heinemann LAJ, Spitzer WO. The differential risk of oral contraceptives: the impact of full exposure history. *Hum Reprod* 1999;14:1493-9.
14 Parkin L, Sharples K, Hernandez RK, Jick SS. Risk of venous thromboembolism in users of oral contraceptives containing drospirenone or levonorgestrel: nested case-control study based on UK General Practice Research Database. *BMJ* 2011;342:d2139.
15 Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to combined oral contraceptives and risk of venous thromboembolism: a protocol for nested case–control studies using the QResearch and the CPRD databases. *BMJ Open* 2014;4:e004499.
16 Stegeman BH, de Bastos M, Rosendaal FR, et al. Different combined oral contraceptives and the risk of venous thrombosis: systematic review and network meta-analysis. *BMJ* 2013;347:f5298.
17 Dunn N. The risk of deep venous thrombosis with oral contraceptives containing drospirenone. *BMJ* 2011;342:d2519.
18 Jick SS, Kaye JA, Vasilakis-Scaramozza C, et al. Validity of the General Practice Research Database. *Pharmacotherapy* 2003;23:686-9.
19 Hippisley-Cox J, Hammersley V, Pringle M, Coupland C, Crown N, Wright L. Methodology for assessing the usefulness of general practice data for research in one research network. *Health Inform J* 2004;10:91-109.
20 Parker C, Coupland C, Hippisley-Cox J. Antipsychotic drugs and risk of venous thromboembolism: nested case-control study. *BMJ* 2010;341:c4245.
21 Vasilakis-Scaramozza C, Jick H. Risk of venous thromboembolism with cyproterone or levonorgestrel contraceptives. *Lancet* 2001;358:1427-29.
22 Etminan M. Pharmacoepidemiology II: the nested case-control study—a novel approach in pharmacoepidemiologic research. *Pharmacotherapy* 2004;24:1105-9.
23 Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions score. *BMJ Open* 2013;3:e003482.
24 Hippisley-Cox J. Validity and completeness of the NHS Number in primary and secondary care electronic data in England 1991-2013. 2013. www.qresearch.org/PowerPointpresentations/Validity%20and%20completeness%20of%20the%20NHS%20Number%20in%20%20primary%20and%20secondary%20care%20data%20in%20England%201991-2013.pdf.
25 Sultan AA, West J, Tata LJ, Fleming KM, Nelson-Piercy C, Grainge MJ. Risk of first venous thromboembolism in and around pregnancy: a population-based cohort study. *Br J Haematol* 2012;156:366-73.
26 World Health Organization Collaborative Study of Cardiovascular Disease Steroid Hormone Contarception. Venous thromboembolic disease and combined oral contraceptives: results of international multicentre case-control study. *Lancet* 1995;346:1575-82.
27 National Institute for Health and Care Excellence. Venous thromboembolic diseases: the management of venous thromboembolic diseases and the role of thrombophilia testing. NICE clinical guideline 144. In: NHS Evidence, 2012.
28 Bird ST, Hartzema AG, Brophy JM, Etminan M, Delaney JAC. Risk of venous thromboembolism in women with polycystic ovary syndrome: a population-based matched cohort analysis. *Can Med Assoc J* 2013;185:E115-20.
29 Lindqvist PG, Epstein E, Olsson H. The relationship between lifestyle factors and venous thromboembolism among women: a report from the MISS study. *Br J Haematol* 2009;144:234-40.
30 French R, Mercer C, Johnson A, Fenton K, Erens B, Wellings K. Use of contraceptive services in Britain: findings from the second National Survey of Sexual Attitudes and Lifestyles (Natsal-2). *J Fam Plann Reprod Health Care* 2009;35:9-14.
31 White RH, Zhou H, Romano PS. Incidence of idiopathic deep venous thrombosis and secondary thromboembolism among ethnic groups in California. *Ann Intern Med* 1998;128:737-40.
32 Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: prospective cohort study. *BMJ* 2011;343:d4656.
33 Royston P. Multiple imputation of missing values. *Stata J* 2004;4:227-41.
34 Borenstein M, Hedges L, Higgins J, Rothstein H. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Syn Meth* 2010;1:97-111.
35 Bjerre L, LeLorier J. Expressing the magnitude of adverse effects in case control studies: "the number needed to be treated for one additional patient to be harmed". *BMJ* 2000;320:503-6.
36 Lader D. Opinions survey report no 41. Contraception and Sexual Health, 2008/09. Office for National Statistics: NHS The Information Centre for health and social care, 2009. www.ons.gov.uk/ons/rel/lifestyles/contraception-and-sexual-health/2008-09/2008-09.pdf.
37 Vaughan B, Trussell J, Kost K, Singh S, Jones R. Discontinuation and resumption of contraceptive use: results from the 2002 National Survey of Family Growth. *Contraception* 2008;78:271-83.
38 The Health and Social Care Information Centre. NHS Contraceptive Services: England, 2004/05. 2005. www.hscic.gov.uk/catalogue/PUB00097/nhs-cont-serv-eng-2004-2005-rep.pdf.
39 Heinemann LAJ, Dinger JC. Range of published estimates of venous thromboembolism incidence in young women. *Contraception* 2007;75:328-36.
40 Lawrenson R, Todd JC, Leydon GM, Williams TJ, Farmer RD. Validation of the diagnosis of venous thromboembolism in general practice database studies. *Br J Clin Pharmacol* 2000;49:591-6.
41 Rothman KJ, Pool C. A strengthening programme for weak associations. *Int J Epidemiol* 1988;17:955-9.
42 Heinemann K, Heinemann LAJ. Comparative risks of venous thromboembolism among users of oral contraceptives containing drospirenone and levonorgestrel. *J Fam Plann Reprod Health Care* 2011;37:132-5.

43 Briggs PE, Praet CA, Humphreys SC, Zhao C. Impact of UK Medical Eligibility Criteria implementation on prescribing of combined hormonal contraceptives. *J Fam Plann Reprod Health Care* 2013;39:190-6.
44 Jick SS, Kaye JA, Russmann S, Jick H. Risk of nonfatal venous thromboembolism with oral contraceptives containing norgestimate or desogestrel compared with oral contraceptives containing levonorgestrel. *Contraception* 2006;73:566-70.
45 Heinemann LAJ, Dinger JC, Assmann A, Minh TD. Use of oral contraceptives containing gestodene and risk of venous thromboembolism: outlook 10 years after the third-generation "pill scare". *Contraception* 2010;81:401-7.
46 Heinemann LAJ, Lewis MA, Assmann A, Thiel C. Case-control studies on venous thromboembolism: bias due to design? A methodological study on venous thromboembolism and steroid hormone use. *Contraception* 2002;65:207-14.
47 Eastwood BJ, Gregor RD, MaClean DR, Wolf HK. Effects of recruitment strategy on response rates and risk factor profile in two cardiovascular surveys. *Int J Epidemiol* 1996;25:763-9.
48 Gronich N, Lavi I, Rennert G. Higher risk of venous thrombosis associated with drospirenone-containing oral contraceptives: a population-based cohort study. *CMAJ* 2011;183:E1319-25.
49 Kuhnz W, Blode H, Mahler M. Systemic availability of levonorgestrel after single oral administration of a norgestimate-containing combination oral contraceptive to 12 young women. *Contraception* 1994;49:255-63.
50 Lidegaard Ø, Milsom IAN, Geirsson RT, Skjeldestad FE. Hormonal contraception and venous thromboembolism. *Acta Obstet Gynecol Scand* 2012;91:769-78.

Accepted: 19 March 2015

Cite this as: *BMJ* 2015;350:h2135

Tables

Table 1| Baseline characteristics in cases and controls by database (CPRD or QResearch)

	CPRD		QResearch	
	Cases (n=5062)	Controls (n=19 638)	Cases (n=5500)	Controls (n=22 396)
Age band at index date				
15-24 years	12.6 (636)	12.7 (2496)	9.0 (493)	9.5 (2135)
25-34 years	25.5 (1290)	23.8 (4666)	25.9 (1423)	25.0 (5589)
35-39 years	17.1 (867)	17.5 (3433)	18.0 (992)	17.7 (3957)
40-44 years	20.8 (1055)	21.9 (4292)	22.5 (1239)	23.3 (5219)
45-49 years	24.0 (1214)	24.2 (4751)	24.6 (1353)	24.5 (5496)
Ethnic group				
White	36.0 (1821)	33.4 (6561)	61.6 (3386)	57.6 (12 900)
Not recorded*	60.2 (3049)	62.4 (12 249)	29.5 (1620)	32.7 (7316)
Black	1.6 (79)	1.2 (237)	4.2 (233)	3.0 (680)
Asian	1.3 (68)	1.9 (375)	2.4 (134)	4.5 (1013)
Other	0.9 (45)	1.1 (216)	2.3 (127)	2.2 (487)
Body mass index				
15-24	34.6 (1753)	44.7 (8774)	34.6 (1903)	44.2 (9895)
25-29	22.6 (1142)	22.0 (4317)	21.9 (1202)	20.0 (4473)
≥30	30.3 (1534)	17.1 (3353)	24.2 (1331)	14.3 (3196)
Not recorded	12.5 (633)	16.3 (3194)	19.3 (1064)	21.6 (4832)
Smoking status				
Non-smoker	51.1 (2586)	54.2 (10 645)	43.5 (2392)	46.5 (10 410)
Ex-smoker	17.5 (884)	16.8 (3295)	23.3 (1280)	22.1 (4952)
Current light smoker	6.3 (319)	6.0 (1188)	14.4 (790)	12.1 (2703)
Current moderate smoker	14.4 (730)	11.2 (2194)	7.7 (424)	6.4 (1433)
Current heavy smoker	6.6 (334)	4.2 (828)	4.5 (248)	2.8 (621)
Not recorded	4.1 (209)	7.6 (1488)	6.7 (366)	10.2 (2277)
Alcohol use				
No use	20.0 (1014)	17.9 (3516)	22.2 (1220)	19.3 (4315)
Ex-use	6.0 (303)	4.4 (869)	6.7 (367)	5.3 (1177)
Light (≤2 units/day)	49.0 (2479)	50.5 (9921)	32.1 (1766)	32.9 (7365)
Moderate/heavy (≥3 units/day)	5.0 (254)	5.0 (986)	17.6 (970)	18.6 (4173)
Not recorded	20.0 (1012)	22.1 (4346)	21.4 (1177)	24.0 (5366)
Non-idiopathic cases				
Proportion (no) of cases or controls	47.0 (2380)	27.2 (5340)	46.9 (2582)	26.3 (5891)
Comorbidities				
Asthma	19.1 (969)	12.9 (2530)	18.8 (1036)	12.0 (2693)
Congestive cardiac disease	0.4 (20)	0.0 (5)	0.2 (13)	0.0 (5)
Rheumatoid arthritis	1.5 (75)	0.6 (121)	2.2 (123)	0.8 (187)
Systemic lupus erythematosus	0.5 (27)	0.1 (22)	0.6 (35)	0.1 (25)
Renal disease	0.9 (48)	0.2 (35)	1.1 (62)	0.3 (65)
Stroke	0.9 (44)	0.1 (22)	0.9 (50)	0.2 (48)
Chronic obstructive pulmonary disease	0.5 (26)	0.2 (30)	0.6 (32)	0.1 (31)
Coronary vascular disease	1.0 (52)	0.3 (50)	1.5 (82)	0.3 (77)
Coagulation disturbances	0.2 (11)	0.0 (9)	0.2 (13)	0.0 (6)
Varicose veins	2.8 (143)	1.6 (314)	2.7 (151)	1.6 (359)
Hypertension	6.3 (319)	3.6 (698)	6.0 (329)	3.7 (831)
Cancer	6.6 (333)	0.9 (180)	6.6 (363)	0.9 (204)

Table 1 (continued)

	CPRD		QResearch	
	Cases (n=5062)	Controls (n=19 638)	Cases (n=5500)	Controls (n=22 396)
Inflammatory bowel disease	1.9 (96)	0.6 (118)	1.8 (100)	0.6 (143)
Conditions in previous 6 months				
Infection	19.0 (964)	10.4 (2033)	17.2 (948)	9.0 (2026)
Surgery or leg/hip fracture	1.1 (54)	0.1 (16)	0.9 (51)	0.1 (24)
Hospital admission	1.4 (72)	0.2 (48)	4.1 (223)	1.1 (248)
Indications for hormonal contraceptive use				
Acne	12.6 (638)	11.7 (2307)	9.3 (514)	8.6 (1933)
Menstrual disorders	36.5 (1847)	31.0 (6091)	27.2 (1497)	23.0 (5141)
Hirsutism	2.1 (107)	1.3 (260)	1.4 (75)	1.0 (229)
Polycystic ovary syndrome	3.4 (174)	2.2 (433)	3.1 (170)	2.4 (535)
Contraceptive drug use in previous month				
Any hormonal contraceptive	32.6 (1649)	20.3 (3996)	33.4 (1838)	19.7 (4418)
Any oral combined contraceptive	24.9 (1259)	14.4 (2835)	23.8 (1309)	12.6 (2823)
Any oral progestogen only	5.1 (260)	4.4 (866)	5.1 (281)	4.0 (907)
Any non-oral hormonal contraceptive	2.6 (130)	1.5 (295)	4.5 (248)	3.1 (688)
Switch in the last month	1.9 (95)	0.6 (110)	1.9 (103)	0.5 (123)

Data are percentage (no) of cases or controls.

*Assumed as white in analyses.

Table 2| **Current exposure to combined oral contraceptives compared to non-exposure by database**

Type of contraceptive	CPRD		QResearch		Combined analysis	
	No of cases/controls	Adjusted odds ratio (95% CI)*	No of cases/controls	Adjusted odds ratio (95% CI)*	Pooled odds ratio (95% CI)	P
Total No	5062/19 638	—	5500/22 396	—	—	—
No use in previous year (reference)	—	1.00	—	1.00	1.00	—
Current use						
Norethisterone	96/245	2.30 (1.78 to 2.99)	109/259	2.82 (2.21 to 3.60)	2.56 (2.15 to 3.06)	<0.001
Levonorgestrel	521/1451	2.23 (1.97 to 2.52)	540/1411	2.52 (2.24 to 2.84)	2.38 (2.18 to 2.59)	<0.001
Norgestimate	122/370	1.96 (1.56 to 2.46)	160/352	3.15 (2.56 to 3.89)	2.53 (2.17 to 2.96)	<0.001
Desogestrel	165/228	4.43 (3.54 to 5.55)	163/262	4.15 (3.34 to 5.15)	4.28 (3.66 to 5.01)	<0.001
Gestodene	78/149	3.14 (2.32 to 4.24)	115/182	4.07 (3.14 to 5.26)	3.64 (3.00 to 4.43)	<0.001
Drospirenone	139/200	4.36 (3.39 to 5.60)	102/170	3.86 (2.93 to 5.08)	4.12 (3.43 to 4.96)	<0.001
Cyproterone	138/192	4.13 (3.22 to 5.31)	120/187	4.42 (3.41 to 5.73)	4.27 (3.57 to 5.11)	<0.001
Different doses of oestrogen						
Norethisterone 20 µg	44/94	2.94 (2.00 to 4.34)	36/79	2.72 (1.78 to 4.16)	2.84 (2.13 to 3.78)	<0.001
Norethisterone 30/40/50 µg	52/151	1.93 (1.36 to 2.72)	73/180	2.87 (2.14 to 3.84)	2.43 (1.94 to 3.03)	<0.001
Desogestrel 20 µg	57/88	4.43 (3.08 to 6.37)	60/97	3.80 (2.68 to 5.41)	4.10 (3.18 to 5.28)	<0.001
Desogestrel 30/40 µg	108/140	4.42 (3.34 to 5.85)	103/165	4.36 (3.33 to 5.71)	4.39 (3.62 to 5.33)	<0.001
Gestodene 20 µg	17/22	4.70 (2.41 to 9.14)	22/25	5.54 (2.99 to 10.28)	5.13 (3.26 to 8.07)	<0.001
Gestodene 30/40 µg	61/127	2.86 (2.05 to 4.00)	93/157	3.83 (2.89 to 5.08)	3.40 (2.74 to 4.21)	<0.001

*Adjusted for body mass index, smoking status, alcohol consumption, ethnic group, chronic and acute conditions, and use of other hormonal contraceptives.

Table 3| **Adjusted odds ratios for current use of different combined oral contraceptives versus levonorgestrel, by database**

Drug name	CPRD		QResearch		Combined analysis	
	Adjusted odds ratio (95% CI)*	P	Adjusted odds ratio (95% CI)*	P	Adjusted odds ratio (95% CI)*	P
Levonorgestrel	1.00	—	1.00	—	1.00	—
Norethisterone	1.03 (0.78 to 1.36)	0.8	1.12 (0.86 to 1.45)	0.4	1.08 (0.89 to 1.30)	0.4
Norgestimate	0.88 (0.69 to 1.12)	0.3	1.25 (1.00 to 1.57)	0.05	1.06 (0.90 to 1.26)	0.5
Desogestrel	1.99 (1.56 to 2.54)	<0.001	1.65 (1.30 to 2.08)	<0.001	1.80 (1.52 to 2.13)	<0.001
Gestodene	1.41 (1.03 to 1.93)	0.03	1.61 (1.23 to 2.12)	<0.001	1.52 (1.24 to 1.87)	<0.001
Drospirenone	1.95 (1.50 to 2.55)	<0.001	1.53 (1.15 to 2.04)	0.004	1.75 (1.43 to 2.12)	<0.001
Cyproterone	1.85 (1.42 to 2.41)	<0.001	1.76 (1.34 to 2.31)	<0.001	1.80 (1.49 to 2.18)	<0.001

*Adjusted for body mass index, smoking status, alcohol consumption, ethnic group, chronic and acute conditions, and use of other hormonal contraceptives.

Table 4| **Additional analyses for current exposure to combined oral contraceptives compared with non-exposure by database**

Type of contraceptive	CPRD		QResearch		Combined analysis	
	No of cases/controls	Adjusted odds ratio (95% CI)*	No of cases/controls	Adjusted odds ratio (95% CI)*	Pooled odds ratio (95% CI)	P
Women treated with anticoagulants						
Total No	2533/9882	—	2956/11 933	—	—	—
No use in previous year	—	1.00	—	1.00	1.00	—
Current use						
Norethisterone	52/131	2.70 (1.88 to 3.87)	57/143	2.82 (2.00 to 3.97)	2.76 (2.16 to 3.54)	<0.001
Levonorgestrel	260/683	2.82 (2.36 to 3.38)	297/739	3.06 (2.59 to 3.61)	2.95 (2.61 to 3.33)	<0.001
Norgestimate	71/181	2.52 (1.84 to 3.46)	99/176	4.68 (3.51 to 6.24)	3.53 (2.86 to 4.37)	<0.001
Desogestrel	113/113	7.37 (5.41 to 10.0)	95/132	5.32 (3.95 to 7.17)	6.23 (5.03 to 7.72)	<0.001
Gestodene	57/61	6.89 (4.56 to 10.4)	82/92	6.20 (4.43 to 8.67)	6.47 (4.98 to 8.39)	<0.001
Drospirenone	94/108	6.03 (4.32 to 8.41)	63/76	6.17 (4.20 to 9.05)	6.09 (4.73 to 7.83)	<0.001
Cyproterone	83/99	5.64 (3.99 to 7.97)	73/95	6.36 (4.45 to 9.08)	5.98 (4.66 to 7.66)	<0.001
Idiopathic cases/controls						
Total No	2630/7632	—	2871/8937	—	—	—
No use in previous year	—	1.00	—	1.00	1.00	—
Current use						
Norethisterone	57/96	2.55 (1.78 to 3.66)	74/117	3.08 (2.24 to 4.24)	2.84 (2.23 to 3.60)	<0.001
Levonorgestrel	321/555	2.70 (2.28 to 3.19)	333/602	2.89 (2.46 to 3.39)	2.80 (2.49 to 3.14)	<0.001
Norgestimate	72/163	1.94 (1.43 to 2.64)	104/148	3.64 (2.74 to 4.82)	2.73 (2.22 to 3.36)	<0.001
Desogestrel	107/100	5.09 (3.75 to 6.91)	98/105	4.73 (3.50 to 6.39)	4.90 (3.95 to 6.08)	<0.001
Gestodene	52/68	3.42 (2.28 to 5.12)	66/72	4.58 (3.20 to 6.58)	4.02 (3.07 to 5.27)	<0.001
Drospirenone	86/78	4.91 (3.44 to 7.01)	68/57	5.61 (3.79 to 8.32)	5.22 (4.01 to 6.79)	<0.001
Cyproterone	83/83	4.77 (3.39 to 6.71)	66/79	4.59 (3.19 to 6.61)	4.69 (3.65 to 6.01)	<0.001
Women aged 15-24 years						
Total No	636/2496	—	493/2135	—	—	—
No use in previous year	—	1.00	—	1.00	1.00	—
Current use						
Norethisterone	15/61	1.10 (0.57 to 2.10)	16/39	3.83 (1.94 to 7.57)	1.99 (1.24 to 3.18)	0.004
Levonorgestrel	150/431	2.42 (1.87 to 3.13)	88/314	2.28 (1.66 to 3.13)	2.36 (1.93 to 2.89)	<0.001
Norgestimate	31/88	2.25 (1.40 to 3.61)	36/76	4.83 (2.97 to 7.84)	3.26 (2.32 to 4.58)	<0.001
Desogestrel	30/49	4.37 (2.57 to 7.44)	24/49	3.52 (1.97 to 6.29)	3.96 (2.67 to 5.86)	<0.001
Gestodene	11/24	2.56 (1.14 to 5.73)	13/25	4.67 (2.21 to 9.88)	3.53 (2.04 to 6.12)	<0.001
Drospirenone	38/64	3.90 (2.37 to 6.40)	17/49	2.69 (1.40 to 5.17)	3.41 (2.29 to 5.05)	<0.001
Cyproterone	37/63	3.77 (2.34 to 6.07)	31/51	4.95 (2.79 to 8.78)	4.21 (2.92 to 6.08)	<0.001
Women aged 25-49 years						
Total No	4426/17142	—	5007/20 261	—	—	—
No use in previous year	—	1.00	—	1.00	1.00	—
Current use						
Norethisterone	81/184	2.75 (2.06 to 3.67)	93/220	2.73 (2.10 to 3.56)	2.74 (2.26 to 3.33)	<0.001
Levonorgestrel	371/1020	2.16 (1.87 to 2.49)	452/1097	2.63 (2.31 to 3.00)	2.40 (2.18 to 2.65)	<0.001
Norgestimate	91/282	1.93 (1.49 to 2.51)	124/276	2.92 (2.31 to 3.70)	2.43 (2.04 to 2.89)	<0.001
Desogestrel	135/179	4.62 (3.59 to 5.93)	139/213	4.26 (3.37 to 5.40)	4.43 (3.73 to 5.26)	<0.001
Gestodene	67/125	3.30 (2.38 to 4.57)	102/157	4.03 (3.06 to 5.30)	3.71 (3.00 to 4.58)	<0.001
Drospirenone	101/136	4.75 (3.53 to 6.38)	85/121	4.37 (3.21 to 5.95)	4.56 (3.69 to 5.65)	<0.001
Cyproterone	101/129	4.41 (3.28 to 5.93)	89/136	4.31 (3.20 to 5.80)	4.36 (3.53 to 5.38)	<0.001

Table 4 (continued)

Type of contraceptive	CPRD		QResearch		Combined analysis	
	No of cases/controls	Adjusted odds ratio (95% CI)*	No of cases/controls	Adjusted odds ratio (95% CI)*	Pooled odds ratio (95% CI)	P

*Adjusted for body mass index, smoking status, alcohol consumption, ethnic group, chronic and acute conditions, and use of other hormonal contraceptives.

Table 5| **Additional analyses for current use of different combined oral contraceptives compared with levonorgestrel by database**

	CPRD		QResearch		Combined analysis	
Drug name	Adjusted odds ratio (95% CI)*	P	Adjusted odds ratio (95% CI)*	P	Adjusted odds ratio (95% CI)*	P
Cases with anticoagulant prescription and matched controls						
Levonorgestrel	1.00	—	1.00	—	1.00	—
Norethisterone	0.96 (0.65 to 1.41)	0.8	0.92 (0.64 to 1.33)	0.7	0.94 (0.72 to 1.22)	0.6
Norgestimate	0.89 (0.64 to 1.26)	0.5	1.53 (1.12 to 2.09)	0.007	1.20 (0.95 to 1.51)	0.1
Desogestrel	2.61 (1.87 to 3.65)	<0.001	1.74 (1.26 to 2.41)	<0.001	2.11 (1.68 to 2.67)	<0.001
Gestodene	2.44 (1.58 to 3.77)	<0.001	2.03 (1.42 to 2.90)	<0.001	2.19 (1.66 to 2.88)	<0.001
Drospirenone	2.14 (1.49 to 3.06)	<0.001	2.02 (1.35 to 3.01)	<0.001	2.08 (1.59 to 2.72)	<0.001
Cyproterone	2.00 (1.38 to 2.89)	<0.001	2.08 (1.43 to 3.03)	<0.001	2.04 (1.57 to 2.65)	<0.001
Idiopathic cases and controls						
Levonorgestrel	1.00	—	1.00	—	1.00	—
Norethisterone	0.94 (0.64 to 1.39)	0.8	1.07 (0.76 to 1.50)	0.7	1.01 (0.78 to 1.30)	0.9
Norgestimate	0.72 (0.52 to 1.00)	0.05	1.26 (0.93 to 1.71)	0.1	0.97 (0.78 to 1.22)	0.8
Desogestrel	1.88 (1.35 to 2.62)	<0.001	1.64 (1.19 to 2.26)	0.003	1.75 (1.39 to 2.21)	<0.001
Gestodene	1.27 (0.83 to 1.94)	0.3	1.59 (1.09 to 2.33)	0.02	1.44 (1.08 to 1.91)	0.01
Drospirenone	1.82 (1.25 to 2.65)	0.002	1.95 (1.29 to 2.94)	0.002	1.88 (1.42 to 2.48)	<0.001
Cyproterone	1.77 (1.23 to 2.53)	0.002	1.59 (1.09 to 2.33)	0.02	1.68 (1.29 to 2.19)	<0.001
Women aged 15-24 years						
Levonorgestrel	1.00	—	1.00	—	1.00	—
Norethisterone	0.45 (0.23 to 0.89)	0.02	1.68 (0.83 to 3.38)	0.1	0.85 (0.52 to 1.38)	0.5
Norgestimate	0.93 (0.57 to 1.52)	0.8	2.12 (1.27 to 3.54)	0.004	1.38 (0.97 to 1.97)	0.08
Desogestrel	1.81 (1.05 to 3.12)	0.03	1.54 (0.85 to 2.81)	0.2	1.68 (1.12 to 2.52)	0.01
Gestodene	1.06 (0.47 to 2.40)	0.9	2.05 (0.94 to 4.44)	0.07	1.50 (0.85 to 2.63)	0.2
Drospirenone	1.61 (0.96 to 2.70)	0.07	1.18 (0.60 to 2.33)	0.6	1.44 (0.95 to 2.17)	0.08
Cyproterone	1.56 (0.95 to 2.56)	0.08	2.17 (1.19 to 3.96)	0.01	1.78 (1.21 to 2.62)	0.003
Women aged 25-49 years						
Levonorgestrel	1.00	—	1.00	—	1.00	—
Norethisterone	1.27 (0.93 to 1.74)	0.1	1.04 (0.78 to 1.38)	0.8	1.14 (0.92 to 1.40)	0.2
Norgestimate	0.89 (0.67 to 1.19)	0.4	1.11 (0.86 to 1.44)	0.4	1.01 (0.83 to 1.22)	0.9
Desogestrel	2.14 (1.62 to 2.82)	<0.001	1.62 (1.25 to 2.10)	<0.001	1.84 (1.53 to 2.23)	<0.001
Gestodene	1.53 (1.08 to 2.16)	0.02	1.53 (1.14 to 2.05)	0.004	1.53 (1.22 to 1.91)	<0.001
Drospirenone	2.20 (1.60 to 3.02)	<0.001	1.66 (1.20 to 2.30)	0.002	1.92 (1.53 to 2.41)	<0.001
Cyproterone	2.04 (1.49 to 2.80)	<0.001	1.64 (1.20 to 2.24)	0.002	1.83 (1.46 to 2.28)	<0.001

*Adjusted for body mass index, smoking status, alcohol consumption, ethnic group, chronic and acute conditions, and use of other hormonal contraceptives.

Table 6| Numbers needed to harm and excess cases per 10 000 patients for different combined oral contraceptives prescribed over one year

	Numbers needed to harm over 1 year (95% CI)		Extra cases per 10 000 treated per year (95% CI)	
Use in previous year	All ages (15-49 years)*	Age 25-49 years†	All ages (15-49 years)*	Age 25-49 years†
Norethisterone	1529 (1159 to 2086)	1169 (874 to 1620)	7 (5 to 9)	9 (6 to 11)
Levonorgestrel	1739 (1506 to 2028)	1452 (1237 to 1723)	6 (5 to 7)	7 (6 to 8)
Norgestimate	1561 (1223 to 2044)	1428 (1077 to 1966)	6 (5 to 8)	7 (5 to 9)
Desogestrel	729 (597 to 899)	594 (478 to 747)	14 (11 to 17)	17 (13 to 21)
Gestodene	905 (697 to 1198)	752 (570 to 1016)	11 (8 to 14)	13 (10 to 18)
Drospirenone	766 (604 to 986)	572 (438 to 758)	13 (10 to 17)	17 (13 to 23)
Cyproterone	731 (582 to 932)	606 (465 to 804)	14 (11 to 17)	17 (12 to 22)

*Based on combined adjusted odds ratios in table 2.

†Based on combined adjusted odds ratios in table 4.

Figures

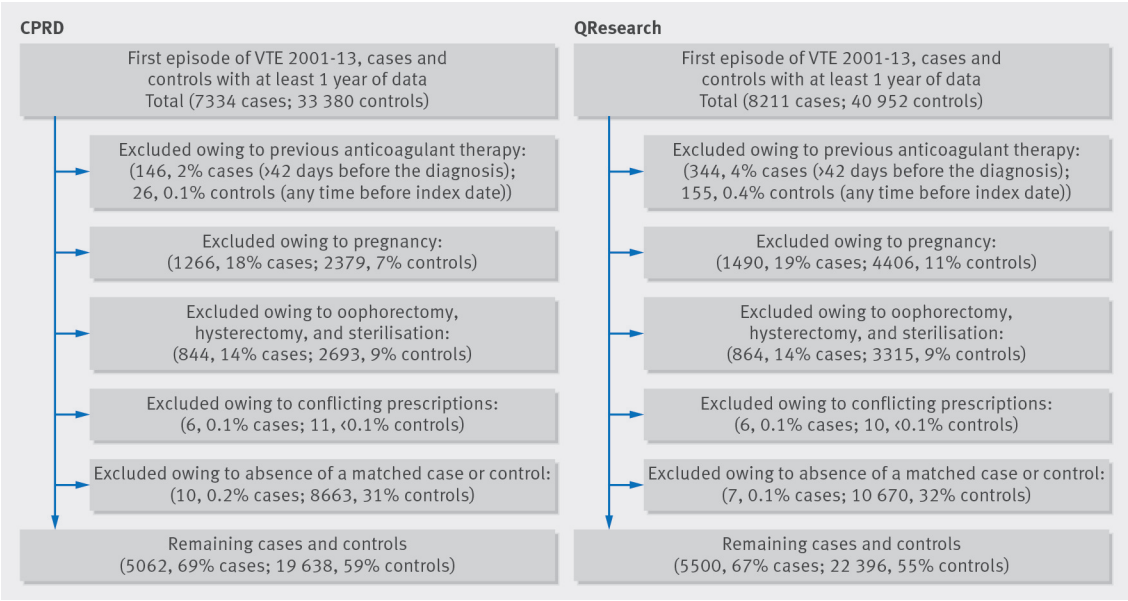

Fig 1 Flow of included patients for CPRD and QResearch analyses with proportions of excluded observations at each point of exclusion

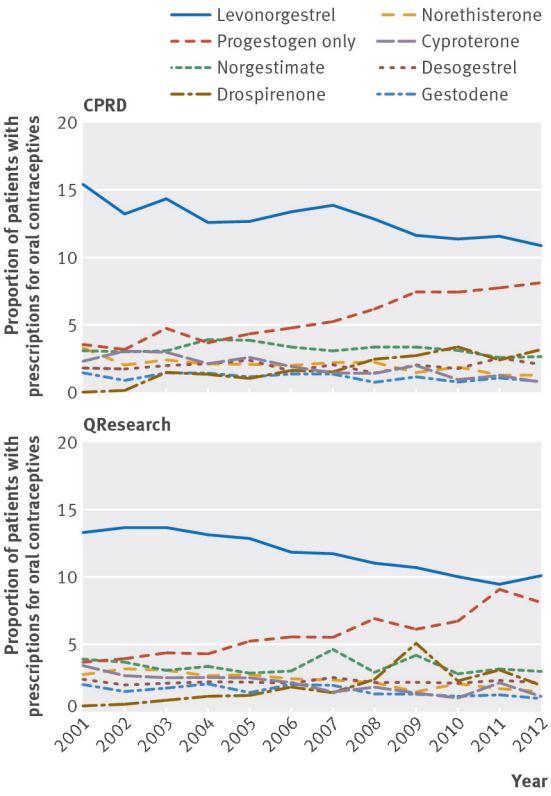

Fig 2 Use of different types of oral contraceptives by year and database. Data are based on age standardised exposure in controls using the UK's general population

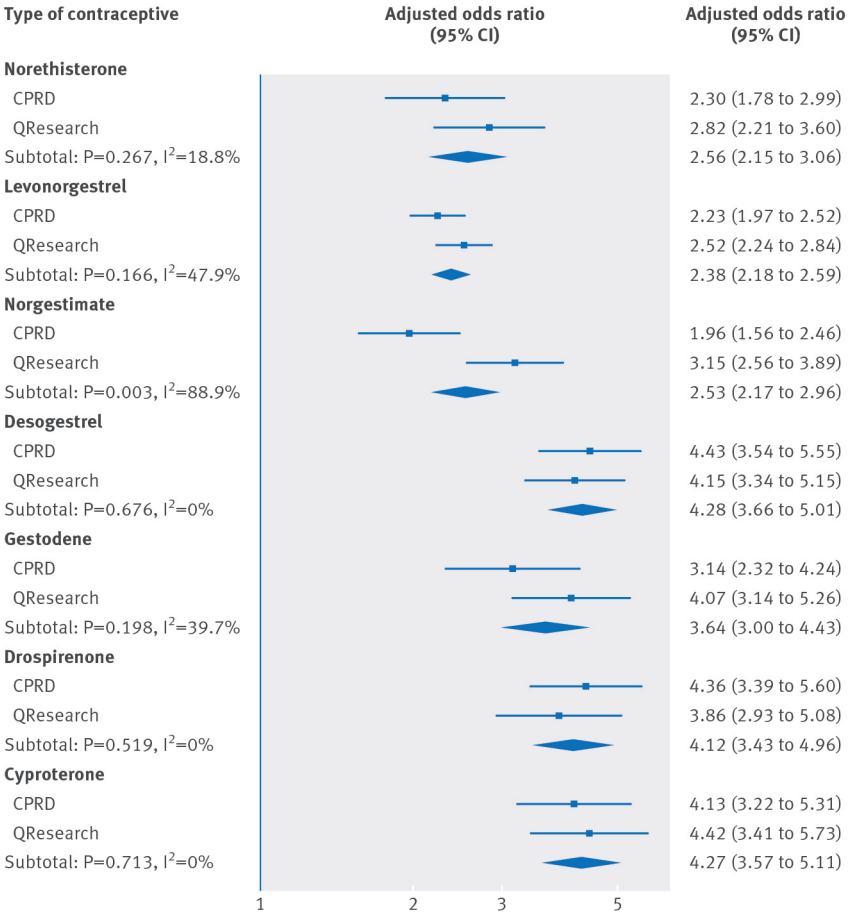

Fig 3 Adjusted odds ratio for VTE in patients currently exposed to combined oral contraceptives compared with no use in the last year, by database. Odds ratios and 95% confidence intervals are adjusted for body mass index, smoking status, alcohol consumption, ethnic group, chronic and acute conditions, and use of other hormonal contraceptives

Paper 9

Hippisley-Cox J, Vinogradova Y, Coupland C, Parker C. Risk of malignancy in patients with schizophrenia or bipolar disorder: nested case-control study. *Archives of General Psychiatry*. 2007;64(12):1368-76.

ORIGINAL ARTICLE

Risk of Malignancy in Patients With Schizophrenia or Bipolar Disorder

Nested Case-Control Study

Julia Hippisley-Cox, MD; Yana Vinogradova, MSc; Carol Coupland, PhD; Chris Parker, MSc

Context: There is conflicting evidence on whether people with schizophrenia have a different risk of cancer from that of the general population.

Objective: To determine the risk of 6 common cancers in patients with schizophrenia or bipolar disorder.

Design: Population-based, nested, case-control study.

Setting: A total of 454 practices contributing to the QRESEARCH general practice database.

Participants: We analyzed 40 441 incident cases of 6 cancers (breast, colon, rectal, gastroesophageal, prostate, and respiratory) and up to 5 controls per case matched by single year of age, sex, general practice, and calendar time.

Main Outcome Measures: Odds ratios (ORs) for cancer risk associated with schizophrenia and bipolar disorder, adjusting for smoking, body mass index, socioeconomic status, comorbidities, and prescribed medications, including antipsychotics.

Results: For breast cancer, we identified 10 535/50 074 cases/controls; colon cancer, 5108/24 458; rectal cancer, 3248/15 552; gastroesophageal cancer, 3854/18 477; prostate cancer, 10 190/48 748; and respiratory cancer, 7506/35 981. After adjustment, patients with schizophrenia had a 190% increased colon cancer risk (adjusted OR, 2.90; 95% confidence interval [CI], 1.85-4.57), a marginal increased breast cancer risk (adjusted OR, 1.52; 95% CI, 1.10-2.11), and a 47% decreased respiratory cancer risk (adjusted OR, 0.53, 95% CI, 0.34-0.85). Patients with schizophrenia taking antipsychotics had a 308% increased colon cancer risk (adjusted OR, 4.08; 95% CI, 2.43-6.84). Patients with bipolar disorder had cancer risks similar to patients with neither condition after adjustment.

Conclusions: Patients with schizophrenia have a significantly higher risk of colon cancer and a lower risk of respiratory cancer compared with patients without schizophrenia after adjustment for confounders. In contrast, the risks of cancer in patients with and without bipolar disorder are similar, suggesting that residual confounding is unlikely to explain the findings. The increased risk of colon cancer is particularly marked in patients with schizophrenia who take antipsychotic medications.

Arch Gen Psychiatry. 2007;64(12):1368-1376

Author Affiliations: Division of Primary Care, University of Nottingham (Drs Hippisley-Cox and Coupland and Ms Vinogradova), and Nottingham Primary Care Research Partnership (Ms Parker), Nottingham, England.

FOR ALMOST 100 YEARS, THERE has been speculation that patients with schizophrenia have lower cancer risks than the general population. In 1909, this possibility was raised by the Board of Control of the Commissioners in Lunacy for England and Wales,[1] but a century later the evidence is still far from clear. Whereas some studies have suggested a lower cancer incidence or mortality rate in people with schizophrenia,[2-7] others have found either an increased cancer incidence[8] or mortality[9] associated with schizophrenia or nonsignificant effects.[10,11] Existing studies have been limited by size, use of biased populations (such as hospital-based cohorts), and lack of ability to control for potential confounding effects. Failure to adjust for these factors is extremely important given that patients with mental health problems have a higher prevalence of common risk factors, including smoking, alcohol consumption, and obesity[12,13]; they are also less likely to report physical symptoms or to adhere to treatment regimens.[14,15]

Considerable uncertainty therefore exists regarding the risk of common cancers in patients with schizophrenia. This is important with respect to designing screening programs as well as etiology. For example, there are theories that schizophrenia itself has a possible protective effect, including a tumor suppressor gene or enhanced natural killer cell activ-

ity.[16,17] Alternatively, medication used in the treatment of schizophrenia may have an antipsychotic effect via the inhibition of enzymes involved in mutation.[18]

Therefore, we undertook a study to compare the risks of 6 common cancers between patients with and without schizophrenia using a very large population-based research database called QRESEARCH, which enabled us to adjust for many potential confounding variables. In addition, we compared cancer risks in patients with bipolar disorder, who have similar lifestyle characteristics as patients with schizophrenia but who are likely to differ with respect to use of medication and any inherent physical correlates of the disease.

METHODS

STUDY POPULATION AND SAMPLE

The full QRESEARCH database (http://www.qresearch.org/) currently contains the anonymized primary care clinical records of more than 10 million people registered at any time in the past 16 years with 525 general practices in the United Kingdom. Consent to provide data is sought from practices using the Egton Medical Information Services (EMIS) medical records system, and detailed analyses have shown that participating practices are somewhat larger than nonparticipating practices but in all other respects are very similar.[19] The database derives from a representative sample of 6% of all the general practices throughout England, Wales, Scotland, and Northern Ireland. The database includes patients' medical records before their registration with any of these practices. The computer system at each participating practice automatically uploads data every 24 hours, ensuring that the most recent information is available. The database has been subjected to detailed analyses of age-sex distributions, birth rates, death rates, consultation rates, prevalence rates, and mortality rates, showing good correspondence with other sources[20] and good levels of completeness and consistency.[21]

We obtained ethical approval from the Trent Multicenter Research Ethics Committee. We used version 7 of the QRESEARCH database, which contained data until August 1, 2005, and included general practices that had used their current computer system for at least 12 months. We identified an open cohort of patients registered with these practices during the 10-year study period (January 1, 1995, to July 1, 2005). For each of 6 common cancers (breast, colon, rectal, gastroesophageal, prostate, and respiratory), we assembled a separate set of individually matched cases and controls from this cohort. Cases all involved patients aged 25 to 100 years with a first-ever record of the index cancer during the study period, including those where the diagnosis was recorded post mortem. We used incidence density sampling to identify up to 5 controls for each incident case matched by single year of age, calendar time, sex, and practice. All the controls were alive and registered with the practice at the date their matched case was first recorded to have the relevant cancer; this was the index date for each case and its controls. Patients were excluded from the study if they had any cancer diagnosis in their record before the first diagnosis of the index cancer (for cases) or the equivalent date (for controls).

We included only patients with at least 12 months of computerized medical record data before their index date to ensure that prescribing data were complete. We excluded breast cancer cases with a record of mastectomy or tamoxifen use more than 12 months before their first record of cancer because these treatments could indicate that they were not incident cases at the time of cancer diagnosis. We also excluded controls with any prior record of mastectomy or tamoxifen use because they could be breast cancer cases without a formal diagnosis in their record.

DATA

We extracted demographic information, including year of birth, sex, and Townsend score (a measure of socioeconomic status). We also extracted each patient's most recent body mass index (BMI) (calculated as weight in kilograms divided by height in meters squared) and smoking status before their index date. Cancer cases and controls were coded as having schizophrenia or bipolar disorder if there was a recorded diagnosis at least 12 months before their index date. We included data on 4 comorbid physical conditions before the index date (ischemic heart disease, diabetes mellitus, hypertension, and rheumatoid arthritis).

We assessed exposure to medications on the basis of at least 1 prescription before the index date. To avoid bias due to reverse causality, we excluded medication used in the 12 months immediately preceding the index date. We included medications for which there was previous evidence of positive or negative association with malignancy, including nonsteroidal anti-inflammatory drugs, cyclooxygenase 2 inhibitors, aspirin, statins, hormone therapy, oral contraceptives, antidepressants (selective serotonin reuptake inhibitors and tricyclic antidepressants), and antipsychotic medications (conventional antipsychotics, atypical antipsychotics, and lithium).

STATISTICAL METHODS

For each cancer, we undertook multiple conditional logistic regression to estimate odds ratios (ORs) associated with schizophrenia and bipolar disorder. We adjusted the models for the possible confounding effects of smoking (current smoker, not current smoker, or not recorded), BMI ($<$25, 25-29, $\geq$30, or not recorded), Townsend score (in fifths), comorbidities (binary), and medications (binary) as well as the other serious mental health condition. We adjusted for use of hormone therapy and oral contraceptives in the breast, colon, and rectal cancer models. All cases and controls were included in the analyses by treating "missing" as a category for the smoking, BMI, and socioeconomic variables. The analyses were repeated on the subset of cases and controls with full data on all variables.

To further examine the relationship between cancer risk and use of antipsychotic medications, we calculated ORs for each cancer for patients with schizophrenia with and without recorded use of antipsychotic medications compared with patients with neither mental health problem. All analyses were conducted using a software program (Stata version 9.1; StataCorp, College Station, Texas).

RESULTS

A total of 454 QRESEARCH practices met the inclusion criteria for this analysis. The total study population consisted of 4 040 494 patients, giving rise to 18 772 868 person-years of observation. We identified 47 924 incident cases of the 6 cancers, of which 7483 met the exclusion criteria, leaving 40 441 cases for analysis. For breast cancer, 740 cases with mastectomy or tamoxifen use more than 12 months previously were excluded, as were 416 controls with any prior record of mastectomy or tamoxifen use. Almost 92% of the respiratory cancer cases were

Table 1. Prevalence of Schizophrenia and Bipolar Disorder in Cases and Controls by Cancer Type

	Cancer					
	Breast	Colon	Rectal	Gastro-esophageal	Prostate	Respiratory
Cancer cases, No. (%)						
Schizophrenia only	47 (0.45)	31 (0.61)	8 (0.25)	12 (0.31)	12 (0.12)	20 (0.27)
Bipolar disorder only	42 (0.40)	12 (0.23)	6 (0.18)	7 (0.18)	15 (0.15)	32 (0.43)
Both conditions	2 (0.02)	1 (0.02)	0	2 (0.05)	1 (0.01)	3 (0.04)
Neither condition	10 444 (99.14)	5064 (99.14)	3234 (99.57)	3833 (99.46)	10 162 (99.73)	7451 (99.27)
Total	**10 535**	**5108**	**3248**	**3854**	**10 190**	**7506**
Cancer controls, No. (%)						
Schizophrenia only	143 (0.29)	48 (0.20)	48 (0.31)	48 (0.26)	108 (0.22)	139 (0.39)
Bipolar disorder only	154 (0.31)	55 (0.22)	30 (0.19)	36 (0.19)	85 (0.17)	94 (0.26)
Both conditions	10 (0.02)	3 (0.01)	2 (0.01)	5 (0.03)	8 (0.02)	9 (0.03)
Neither condition	49 767 (99.39)	24 352 (99.57)	15 472 (99.49)	18 388 (99.52)	48 547 (99.59)	35 739 (99.33)
Total	**50 074**	**24 458**	**15 552**	**18 477**	**48 748**	**35 981**

Table 2. Characteristics of Cases and Controls by Cancer Type

	Breast Cancer		Colon Cancer		Rectal Cancer	
Characteristic	Cases (n = 10 535)	Controls (n = 50 074)	Cases (n = 5108)	Controls (n = 24 458)	Cases (n = 3248)	Controls (n = 15 552)
Sex, No. (%)						
M	10 (0.1)	46 (0.1)	2669 (52.3)	12 758 (52.2)	2013 (62.0)	9617 (61.8)
F	10 525 (99.9)	50 028 (99.9)	2439 (47.7)	11 700 (47.8)	1235 (38.0)	5935 (38.2)
Age, median (interquartile range), y	61 (51 to 72)	61 (51 to 72)	72 (64 to 79)	72 (64 to 79)	71 (62 to 78)	71 (62 to 78)
Townsend score recorded, No. (%)	10 335 (98.1)	48 578 (97.0)	4959 (97.1)	23 557 (96.3)	3150 (97.0)	14 980 (96.3)
Townsend score, median (interquartile range)	–1.61 (–3.24 to 1.25)	–1.48 (–3.22 to 1.33)	–1.27 (–3.07 to 1.53)	–1.41 (–3.17 to 1.50)	–1.06 (–3.01 to 1.83)	–1.21 (–3.06 to 1.81)
Smoking status recorded, No. (%)	8960 (85.0)	40 784 (81.4)	4258 (83.4)	19 261 (78.8)	2653 (81.7)	12 217 (78.6)
Smokers, No. (%)	2073 (19.7)	9641 (19.3)	803 (15.7)	3911 (16.0)	664 (20.4)	2727 (17.5)
BMI recorded, No. (%)	8147 (77.3)	37 071 (74.0)	3831 (75.0)	17 196 (70.3)	2359 (72.6)	10 834 (69.7)
BMI, median (interquartile range)	25.5 (22.8 to 29.0)	25.4 (22.7 to 29.1)	26.1 (23.5 to 29.0)	26.0 (23.5 to 29.0)	26.1 (23.5 to 28.9)	26.1 (23.6 to 29.0)
Months of previous data, median (interquartile range)	61 (37 to 95)	61 (37 to 94)	66 (40 to 102)	67 (40 to 102)	68 (39 to 102)	69 (39 to 102)
Comorbidities, No. (%)						
Diabetes mellitus	431 (4.1)	2028 (4.1)	424 (8.3)	1553 (6.3)	254 (7.8)	1021 (6.6)
Ischemic heart disease	546 (5.2)	2793 (5.6)	683 (13.4)	3092 (12.6)	404 (12.4)	2057 (13.2)
Rheumatoid arthritis	131 (1.2)	745 (1.5)	40 (0.8)	339 (1.4)	39 (1.2)	213 (1.4)
Hypertension	2292 (21.8)	10 520 (21.0)	1452 (28.4)	6877 (28.1)	927 (28.5)	4080 (26.2)
Use of medications, No. (%)						
NSAIDs	1273 (12.1)	5832 (11.6)	562 (11.0)	2916 (11.9)	362 (11.1)	1934 (12.4)
Aspirin	1270 (12.1)	5830 (11.6)	1139 (22.3)	5295 (21.6)	667 (20.5)	3342 (21.5)
Cyclooxygenase 2 inhibitors	591 (5.6)	2442 (4.9)	240 (4.7)	1075 (4.4)	111 (3.4)	681 (4.4)
Statins	573 (5.4)	2855 (5.7)	454 (8.9)	2149 (8.8)	271 (8.3)	1339 (8.6)
Hormone therapy	2458 (23.3)	9940 (19.9)	283 (5.5)	1363 (5.6)	140 (4.3)	765 (4.9)
Oral contraceptives	626 (5.9)	2573 (5.1)	35 (0.7)	151 (0.6)	24 (0.7)	104 (0.7)
Antidepressant SSRIs	1219 (11.6)	5736 (11.5)	374 (7.3)	1747 (7.1)	195 (6.0)	1095 (7.0)
Antidepressant TCAs	1855 (17.6)	8344 (16.7)	682 (13.4)	3196 (13.1)	367 (11.3)	2034 (13.1)
Antipsychotics	1478 (14.0)	6643 (13.3)	639 (12.5)	2962 (12.1)	334 (10.3)	1806 (11.6)

(continued)

lung cancer (6894 of 7506). **Table 1** gives the number of cases and controls for each cancer and the prevalence of schizophrenia and bipolar disorder in each group. Nine cases and 37 controls had diagnoses of both schizophrenia and bipolar disorder. **Table 2** summarizes the characteristics of cases and controls for each cancer, and **Table 3** compares patients with and without schizophrenia or bipolar disorder, showing that the groups with mental health problems were somewhat younger and more likely to smoke and had lower rates of some comorbidities than the group with neither mental health problem.

Table 4 gives the ORs for each cancer associated with schizophrenia and bipolar disorder, unadjusted and adjusted for socioeconomic status, smoking, BMI, comorbidities, and use of medications. The adjusted analysis (also shown in the **Figure**) shows that people with schizo-

Table 2. Characteristics of Cases and Controls by Cancer Type (cont)

Characteristic	Gastroesophageal Cancer		Prostate Cancer		Respiratory Cancer	
	Cases (n = 3854)	Controls (n = 18 477)	Cases (n = 10 190)	Controls (n = 48 748)	Cases (n = 7506)	Controls (n = 35 981)
Sex, No. (%)						
M	2503 (64.9)	11 966 (64.8)	10 190 (100)	48 748 (100)	4843 (64.5)	23 156 (64.4)
F	1351 (35.1)	6511 (35.2)	0	0	2663 (35.5)	12 825 (35.6)
Age, median (interquartile range), y	72 (64 to 79)	72 (64 to 79)	73 (67 to 79)	73 (67 to 79)	71 (63 to 78)	71 (63 to 77)
Townsend score recorded, No. (%)	3743 (97.1)	17 850 (96.6)	9989 (98.0)	47 222 (96.9)	7305 (97.3)	34 729 (96.5)
Townsend score, median (interquartile range)	–1.05 (–3.01 to 2.05)	–1.34 (–3.14 to 1.76)	–1.80 (–3.38 to 0.99)	–1.60 (–3.26 to 1.19)	–0.01 (–2.48 to 3.03)	–0.92 (–2.92 to 2.26)
Smoking status recorded, No. (%)	3251 (84.4)	14 674 (79.4)	8727 (85.6)	38 499 (79.0)	6528 (87.0)	28 965 (80.5)
Smokers, No. (%)	1007 (26.1)	3230 (17.5)	1612 (15.8)	8149 (16.7)	3554 (47.3)	6747 (18.8)
BMI recorded, No. (%)	2884 (74.8)	13 008 (70.4)	7893 (77.5)	34 342 (70.4)	5637 (75.1)	25 820 (71.8)
BMI, median (interquartile range)	25.5 (22.9 to 28.6)	26.0 (23.5 to 28.9)	26.0 (23.9 to 28.4)	26.1 (23.8 to 28.7)	25.0 (22.4 to 28.1)	26.1 (23.6 to 29.0)
Months of previous data, median (interquartile range)	69 (42 to 103)	69 (42 to 102)	71 (42 to 107)	72 (42 to 106)	66 (39 to 100)	65 (39 to 99)
Comorbidities, No. (%)						
Diabetes mellitus	340 (8.8)	1240 (6.7)	750 (7.4)	3835 (7.9)	534 (7.1)	2499 (6.9)
Ischemic heart disease	561 (14.6)	2585 (14.0)	1714 (16.8)	8076 (16.6)	1161 (15.5)	5115 (14.2)
Rheumatoid arthritis	40 (1.0)	264 (1.4)	92 (0.9)	460 (0.9)	143 (1.9)	477 (1.3)
Hypertension	1058 (27.5)	4960 (26.8)	2966 (29.1)	13 094 (26.9)	1898 (25.3)	9953 (27.7)
Use of medications, No. (%)						
NSAIDs	476 (12.4)	2195 (11.9)	1273 (12.5)	5535 (11.4)	880 (11.7)	4120 (11.5)
Aspirin	919 (23.8)	4160 (22.5)	2730 (26.8)	12 510 (25.7)	1980 (26.4)	8204 (22.8)
Cyclooxygenase 2 inhibitors	174 (4.5)	788 (4.3)	472 (4.6)	1935 (4.0)	390 (5.2)	1785 (5.0)
Statins	344 (8.9)	1584 (8.6)	1159 (11.4)	5121 (10.5)	787 (10.5)	3709 (10.3)
Hormone therapy	136 (3.5)	663 (3.6)	0	0	433 (5.8)	1862 (5.2)
Oral contraceptives	24 (0.6)	128 (0.7)	0	0	33 (0.4)	167 (0.5)
Antidepressant SSRIs	257 (6.7)	1220 (6.6)	561 (5.5)	2554 (5.2)	703 (9.4)	2674 (7.4)
Antidepressant TCAs	541 (14.0)	2271 (12.3)	1028 (10.1)	4605 (9.4)	1261 (16.8)	4613 (12.8)
Antipsychotics	486 (12.6)	2196 (11.9)	988 (9.7)	4678 (9.6)	940 (12.5)	4236 (11.8)

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); NSAIDs, nonsteroidal anti-inflammatory drugs; SSRIs, selective serotonin reuptake inhibitors; TCAs, tricyclic antidepressants.

phrenia had significantly higher risks of breast cancer (adjusted OR, 1.52; 95% confidence interval [CI], 1.10-2.11) and colon cancer (adjusted OR, 2.90; 95% CI, 1.85-4.57) and a significantly lower risk of respiratory cancer (adjusted OR, 0.53; 95% CI, 0.34-0.85). There were no significant differences in their risks of rectal, gastroesophageal, or prostate cancer. Patients with bipolar disorder had ORs closer to unity for each cancer, and there were no statistically significant associations after adjustment for confounders.

Restricting the respiratory cancer analysis to lung cancer cases and their controls made little difference to the observed reduction in risk for people with schizophrenia (adjusted OR, 0.59; 95% CI, 0.37-0.95). Restricting the analysis for each cancer to patients with full data on all confounding variables made no substantial difference to the adjusted ORs for colon, rectal, gastroesophageal, or prostate cancer. For breast and respiratory cancers, the adjusted ORs were closer to unity and the CIs were wider (breast: OR, 1.28; 95% CI, 0.84-1.96; respiratory: OR, 0.67; 95% CI, 0.39-1.15).

Overall, 489 of the 710 patients with schizophrenia (68.9%) had 1 or more prescriptions for antipsychotic medications at least 12 months before their index date. **Table 5** shows separate ORs for each cancer for patients with schizophrenia who had or had not been prescribed antipsychotic medications compared with patients with neither mental health condition, first unadjusted and then adjusted as described in the "Statistical Methods" subsection of the "Methods" section. The 52.2% increase in breast cancer risk found for patients with schizophrenia overall (after adjustment) was not substantially different in the subgroups with and without antipsychotic medication use (55.0% and 42.6%, respectively). The almost 3-fold increased risk of colon cancer in patients with schizophrenia overall was stronger (>4-fold) in the subgroup of patients with schizophrenia also prescribed antipsychotic medications. The overall 46.5% reduced risk of respiratory cancer in patients with schizophrenia was most marked (85.8%) in the subgroup not taking antipsychotics. In view of the small numbers in some of the subgroup analyses, particularly for rectal and gastroesophageal cancers, caution is needed in interpreting these ORs.

COMMENT

This is a very large population-based study to determine the risks of 6 common cancers (breast, colon, rectal, gastroesophageal, prostate, and respiratory) in patients with schizophrenia or bipolar disorder. The key finding

Table 3. Characteristics of Patients With and Without Schizophrenia or Bipolar Disorder

Characteristic	Patients With Schizophrenia[a]		Patients With Bipolar Disorder[a]		Patients With Neither Condition	
	Cases (n = 139)	Controls (n = 571)	Cases (n = 123)	Controls (n = 491)	Cases (n = 40 188)	Controls (n = 192 265)
Sex, No. (%)						
M	56 (40.3)	289 (50.6)	50 (40.7)	212 (43.2)	22 125 (55.1)	105 808 (55.0)
F	83 (59.7)	282 (49.4)	73 (59.3)	279 (56.8)	18 063 (44.9)	86 457 (45.0)
Age, median (interquartile range), y	67 (58 to 74)	68 (59 to 75)	67 (60 to 75)	69 (60 to 76)	70 (61 to 78)	70 (61 to 78)
Townsend score recorded, No. (%)	135 (97.1)	554 (97.0)	121 (98.4)	461 (93.9)	39 242 (97.6)	185 937 (96.7)
Townsend score, median (interquartile range)	1.16 (–1.98 to 4.02)	1.26 (–1.81 to 4.35)	–0.56 (–2.53 to 3.60)	–0.55 (–2.84 to 2.68)	–1.28 (–3.10 to 1.68)	–1.38 (–3.16 to 1.55)
Smoking status recorded, No. (%)	120 (86.3)	436 (76.4)	107 (87.0)	418 (85.1)	34 159 (85.0)	153 574 (79.9)
Smokers, No. (%)	59 (42.4)	185 (32.4)	56 (45.5)	135 (27.5)	9603 (23.9)	34 101 (17.7)
BMI recorded, No. (%)	103 (74.1)	372 (65.1)	99 (80.5)	379 (77.2)	30 558 (76.0)	137 545 (71.5)
BMI, median (interquartile range)	26.5 (23.3 to 30.8)	26.0 (23.0 to 29.9)	25.1 (22.1 to 29.1)	26.3 (23.1 to 29.8)	25.7 (23.1 to 28.7)	25.9 (23.4 to 28.9)
Months of previous data, median (interquartile range)	60 (40 to 93)	58 (35 to 88)	61 (37 to 96)	59 (35 to 98)	67 (39 to 101)	67 (39 to 101)
Comorbidities, No. (%)						
Diabetes mellitus	9 (6.5)	47 (8.2)	10 (8.1)	45 (9.2)	2715 (6.8)	12 089 (6.3)
Ischemic heart disease	10 (7.2)	47 (8.2)	15 (12.2)	79 (16.1)	5045 (12.6)	23 598 (12.3)
Rheumatoid arthritis	3 (2.2)	2 (0.4)	1 (0.8)	6 (1.2)	481 (1.2)	2490 (1.3)
Hypertension	26 (18.7)	94 (16.5)	28 (22.8)	90 (18.3)	10 543 (26.2)	49 300 (25.6)
Use of medications, No. (%)						
NSAIDs	17 (12.2)	25 (4.4)	15 (12.2)	58 (11.8)	4795 (11.9)	22 452 (11.7)
Aspirin	23 (16.5)	97 (17.0)	29 (23.6)	107 (21.8)	8655 (21.5)	39 144 (20.4)
Cyclooxygenase 2 inhibitors	5 (3.6)	15 (2.6)	12 (9.8)	23 (4.7)	1961 (4.9)	8671 (4.5)
Statins	7 (5.0)	31 (5.4)	12 (9.8)	41 (8.4)	3570 (8.9)	16 688 (8.7)
Hormone therapy	11 (7.9)	22 (3.9)	24 (19.5)	72 (14.7)	3419 (8.5)	14 511 (7.5)
Oral contraceptives	0	2 (0.4)	1 (0.8)	6 (1.2)	744 (1.9)	3120 (1.6)
Antidepressant SSRIs	29 (20.9)	100 (17.5)	57 (46.3)	151 (30.8)	3226 (8.0)	14 783 (7.7)
Antidepressant TCAs	39 (28.1)	114 (20.0)	54 (43.9)	241 (49.1)	5645 (14.0)	24 725 (12.9)
Antipsychotics	110 (79.1)	379 (66.4)	65 (52.8)	253 (51.5)	4698 (11.7)	21 918 (11.4)

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); NSAIDs, nonsteroidal anti-inflammatory drugs; SSRIs, selective serotonin reuptake inhibitors; TCAs, tricyclic antidepressants.

[a] Forty-six patients had both schizophrenia and bipolar disorder and are included in both groups.

in patients with schizophrenia was a 3-fold increased risk of colon cancer, which was more than 4-fold in patients also prescribed antipsychotic medications, despite adjustment for the potential confounding effects of socioeconomic status, smoking, obesity, comorbidity, and concurrent use of other medications. This is a novel and unexpected finding that needs further investigation. We also found a 52.2% increased risk of breast cancer, and a 46.5% decreased risk of respiratory cancer (predominantly lung cancer) after adjustment for the same variables. Patients with bipolar disorder had similar cancer risks as people without either mental health condition after adjustment for the potential confounding variables.

COLON AND RECTAL CANCERS

The most important finding in this study is a marked increased risk of colon cancer in patients with schizophrenia. This is a novel finding and one that is in contrast with previous studies,[4,8,22] which have tended to suggest no difference in risk of colon cancer in patients with schizophrenia. The previous UK study by Goldacre et al[22] found a nonsignificant rate ratio of 0.72. Their study differed from this one in a variety of ways, being based on records of hospital admissions rather than on the primary care records of the whole population, the reference cohort being drawn from patients admitted to the hospital for other conditions, and rates being standardized by age and sex but no adjustment being made for potential confounding variables. The finding of an almost 3-fold increased risk of colon cancer persisted whether or not the results were adjusted for known risk factors, including obesity[23] and socioeconomic status,[24] and for use of hormone therapy or oral contraceptives, which might reduce risk.[25] We did not adjust for diet, exercise, or alcohol consumption,[26] and these factors tend not to be reliably recorded in electronic medical records. We found that the increase in colon cancer was greatest in patients with schizophrenia prescribed antipsychotic agents. This is in contrast to a recent study[27] that reported a reduction in risk associated with these drugs,

Table 4. Odds Ratios for Risk of Each Cancer Associated With Schizophrenia and Bipolar Disorder[a]

	Participants, No. (%)		Odds Ratio (95% CI)[b]	
	Cases	Controls	Unadjusted	Adjusted
Breast cancer	n = 10 535	n = 50 074		
Neither mental health problem	10 444 (99.14)	49 767 (99.39)	1 [Reference]	1 [Reference]
Schizophrenia	49 (0.47)	153 (0.31)	1.50 (1.08-2.07)	1.52 (1.10-2.11)
Bipolar disorder	44 (0.42)	164 (0.33)	1.25 (0.89-1.76)	1.21 (0.86-1.71)
Colon cancer	n = 5108	n = 24 458		
Neither mental health problem	5064 (99.14)	24 352 (99.57)	1 [Reference]	1 [Reference]
Schizophrenia	32 (0.63)	51 (0.21)	2.85 (1.82-4.45)	2.90 (1.85-4.57)
Bipolar disorder	13 (0.25)	58 (0.24)	0.97 (0.52-1.79)	0.95 (0.51-1.76)
Rectal cancer	n = 3248	n = 15 552		
Neither mental health problem	3234 (99.57)	15 472 (99.49)	1 [Reference]	1 [Reference]
Schizophrenia	8 (0.25)	50 (0.32)	0.77 (0.36-1.63)	0.78 (0.36-1.66)
Bipolar disorder	6 (0.18)	32 (0.21)	0.91 (0.37-2.24)	0.99 (0.40-2.43)
Gastroesophageal cancer	n = 3854	n = 18 477		
Neither mental health problem	3833 (99.46)	18 388 (99.52)	1 [Reference]	1 [Reference]
Schizophrenia	14 (0.36)	53 (0.29)	1.25 (0.69-2.27)	1.06 (0.58-1.93)
Bipolar disorder	9 (0.23)	41 (0.22)	1.05 (0.50-2.17)	0.98 (0.47-2.05)
Prostate cancer	n = 10 190	n = 48 748		
Neither mental health problem	10 162 (99.73)	48 547 (99.59)	1 [Reference]	1 [Reference]
Schizophrenia	13 (0.13)	116 (0.24)	0.54 (0.30-0.95)	0.59 (0.33-1.05)
Bipolar disorder	16 (0.16)	93 (0.19)	0.86 (0.50-1.47)	0.87 (0.51-1.49)
Respiratory cancer	n = 7506	n = 35 981		
Neither mental health problem	7451 (99.27)	35 739 (99.33)	1 [Reference]	1 [Reference]
Schizophrenia	23 (0.31)	148 (0.41)	0.71 (0.46-1.11)	0.53 (0.34-0.85)
Bipolar disorder	35 (0.47)	103 (0.29)	1.68 (1.13-2.48)	1.21 (0.79-1.85)

Abbreviation: CI, confidence interval.
[a] Mental health categories do not sum to overall totals because 46 patients had both schizophrenia and bipolar disorder and are included in both groups.
[b] All models are adjusted for smoking, obesity, socioeconomic status, diabetes mellitus, hypertension, ischemic heart disease, rheumatoid arthritis, and use of nonsteroidal anti-inflammatory drugs, cyclooxygenase 2 inhibitors, aspirin, statins, selective serotonin reuptake inhibitors, tricyclic antidepressants, and antipsychotics. In addition, the breast, colon, and rectal cancer models are adjusted for use of hormone therapy and oral contraceptives.

although this was found in a wider population, of whom only 6% had schizophrenia.

For rectal cancer, which has few established risk factors, we found no significant increase or decrease in risk for patients with severe mental illness, in contrast to other studies,[4,8,22] which demonstrate a decrease in risk, but our numbers in these subgroups were small.

BREAST CANCER

We found a marginal 52.2% increase in risk of breast cancer in women with schizophrenia. This is consistent with some[4,28,29] but not all studies, which have generally found no difference[5,8,22,30] or a reduction in risk.[2] There have been suggestions that an increase in risk of breast cancer could be mediated by a prolactin-releasing effect of neuroleptic medications,[31] but a recent study did not confirm this,[27] and we found only a marginal association with medication. We adjusted for some risk factors that could confound the relationship between schizophrenia and cancer risk, including obesity[32] and use of oral contraceptives or hormone therapy.[33,34] However, we did not adjust for the observed lower parity in women with schizophrenia,[4] which tends to be associated with an increased risk of breast cancer. In other words, it is possible that the marginal increased risk of breast cancer demonstrated in this study is due to residual confounding by lower parity rather than a true increase in risk.

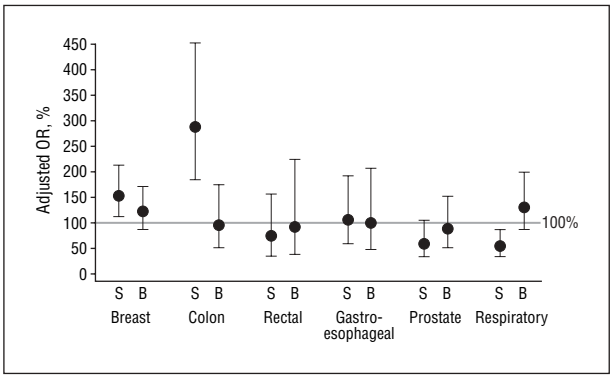

Figure. Adjusted odds ratios (ORs) for each cancer for patients with schizophrenia (S) and bipolar disorder (B) compared with patients without mental health problems. Error bars represent 95% confidence intervals.

RESPIRATORY CANCER

Some previous studies have reported a higher incidence or mortality due to respiratory cancer,[6,8,10] but others report a lower risk consistent with the present findings[4,7] or a nonsignificant difference.[5,22] Smoking is a major risk factor for lung cancer and is more prevalent in people with schizophrenia,[12] making it a potential source of substantial confounding in these studies; it was adjusted for in the present analysis. The largely institutionalized life of many people with schizophrenia could protect them

Table 5. Odds Ratios for Risk of Each Cancer in Patients With Schizophrenia With and Without Antipsychotic Medication Use

	Participants, No. (%)		Odds Ratio (95% CI)	
	Cases	Controls	Unadjusted	Adjusted[a]
Breast cancer	n = 10 535	n = 50 074		
Neither mental health problem	10 444 (99.14)	49 767 (99.39)	1 [Reference]	1 [Reference]
Schizophrenia without medication	9 (0.09)	31 (0.06)	1.34 (0.64-2.83)	1.43 (0.68-3.01)
Schizophrenia with medication	40 (0.38)	122 (0.24)	1.51 (1.05-2.16)	1.55 (1.08-2.23)
Colon cancer	n = 5108	n = 24 458		
Neither mental health problem	5064 (99.14)	24 352 (99.57)	1 [Reference]	1 [Reference]
Schizophrenia without medication	4 (0.08)	19 (0.08)	0.92 (0.31-2.74)	0.97 (0.32-2.91)
Schizophrenia with medication	28 (0.55)	32 (0.13)	4.07 (2.43-6.82)	4.08 (2.43-6.84)
Rectal cancer	n = 3248	n = 15 552		
Neither mental health problem	3234 (99.57)	15 472 (99.49)	1 [Reference]	1 [Reference]
Schizophrenia without medication	0	20 (0.13)	NA	NA
Schizophrenia with medication	8 (0.25)	30 (0.19)	1.29 (0.59-2.81)	1.26 (0.58-2.78)
Gastroesophageal cancer	n = 3854	n = 18 477		
Neither mental health problem	3833 (99.46)	18 388 (99.52)	1 [Reference]	1 [Reference]
Schizophrenia without medication	7 (0.18)	24 (0.13)	1.34 (0.57-3.14)	1.07 (0.45-2.53)
Schizophrenia with medication	7 (0.18)	29 (0.16)	1.19 (0.52-2.74)	1.07 (0.46-2.49)
Prostate cancer	n = 10 190	n = 48 748		
Neither mental health problem	10 162 (99.73)	48 547 (99.59)	1 [Reference]	1 [Reference]
Schizophrenia without medication	7 (0.07)	47 (0.10)	0.69 (0.31-1.54)	0.75 (0.34-1.68)
Schizophrenia with medication	6 (0.06)	69 (0.14)	0.43 (0.18-0.98)	0.47 (0.20-1.08)
Respiratory cancer	n = 7506	n = 35 981		
Neither mental health problem	7451 (99.27)	35 739 (99.33)	1 [Reference]	1 [Reference]
Schizophrenia without medication	2 (0.03)	51 (0.14)	0.18 (0.04-0.76)	0.14 (0.03-0.60)
Schizophrenia with medication	21 (0.28)	97 (0.27)	0.98 (0.61-1.58)	0.72 (0.44-1.18)

Abbreviations: CI, confidence interval; NA, not applicable.

[a] All models are adjusted for smoking, obesity, socioeconomic status, diabetes mellitus, hypertension, ischemic heart disease, rheumatoid arthritis, bipolar disorder, and use of nonsteroidal anti-inflammatory drugs, cyclooxygenase 2 inhibitors, aspirin, statins, selective serotonin reuptake inhibitors, and tricyclic antidepressants. In addition, the breast, colon, and rectal cancer models are adjusted for use of hormone therapy and oral contraceptives.

from some environmental risks, and antipsychotic drugs have been suggested as being protective, but in this sample the reduction in risk was greatest in patients not taking antipsychotic medications. The authors of a recent study[27] that found an increased risk of lung cancer associated with antipsychotic medication suggested residual confounding by smoking. Again, we found that patients with bipolar disorder did not share the reduced risk associated with schizophrenia.

GASTROESOPHAGEAL CANCER

We found no significant difference in risk of gastroesophageal cancer in people with schizophrenia, which is consistent with 2 previous studies.[2,8] Other studies have reported a higher risk of cancer of the esophagus after adjusting for age and sex[4,22] and reduced mortality from gastric cancer.[35] In the present study, adjustment for potential confounders, including socioeconomic status,[24] moved the OR closer to unity, but we did not adjust for alcohol consumption, which is a strong risk factor for cancers of the upper digestive tract.[36]

PROSTATE CANCER

The adjusted analysis suggests a 40.7% lower risk of prostate cancer in people with schizophrenia, but in view of the small numbers in these subgroups, this was not a statistically significant reduction. The magnitude of the association is consistent with that of previous studies, which have shown an approximately 50% lower risk.[4,6-8] Little is known about risk factors for prostate cancer, but a protective effect of neuroleptic medications, particularly phenothiazines, has been suggested[27,37]; our subgroup analysis is consistent with this but is based on numbers too small to reach a firm conclusion. Alternatively, the lower risk could represent an ascertainment bias if patients with schizophrenia are less likely to have screening for prostate cancer.

STRENGTHS AND LIMITATIONS OF THIS STUDY

Aggregated general practice databases, such as QRESEARCH, have previously been used successfully to evaluate risk factors for diseases in the population.[38-41] The use of routine clinical records and a nested case-control design gave this study a large and representative population-based sample, matched cancer cases and controls, no response or recall bias, and a comparison group with another mental health condition (bipolar disorder). Recording of clinical diagnoses was shown to have good levels of accuracy and completeness in general practice in the United Kingdom, including malignancy and psychiatric illness, where the diagnosis is recorded after specialist investigations and consultations.[42,43] The quality of the electronic medical record is thought to be highest in practices that contribute to primary care data-

bases.[44] Advantages of this study in contrast to previous work are that we adjusted for confounders such as smoking, BMI, and socioeconomic status and for commonly used medications, including antipsychotics. Information on alcohol consumption, diet, exercise, and reproductive history is less reliably recorded and was not included. However, in people with bipolar disorder, who would be expected to share some of the increased risk attributable to these lifestyle factors, we did not find the associations with cancer observed in patients with schizophrenia. Even with almost 19 million person-years of observation, the small number of patients with mental health problems in some subgroups limited the size of the effect that could be detected.

It is possible that some cancers were undiagnosed in people with schizophrenia or bipolar disorder, perhaps owing to underreporting of physical problems or lack of participation in screening programs. However, such a misclassification of outcomes would tend to bias the OR downward rather than generating spuriously positive findings, such as the increased risk of colon cancer in patients with schizophrenia. By excluding diagnoses of schizophrenia or bipolar disorder made in the 12 months preceding the diagnosis of cancer, we minimized the possibility that apparent psychiatric symptoms were a manifestation of the cancer.

IMPLICATIONS FOR RESEARCH AND CLINICAL PRACTICE

The possible association between schizophrenia and increased risk of colon and breast cancers is of practical and theoretical importance not only in terms of the organization of services (such as screening) but also in the understanding of the etiology of disease. Given the study design, it is not possible to eliminate the possibility of residual confounding by such factors as alcohol consumption, diet, and reproductive history. However, in people with bipolar disorder, who would be expected to share some of the increased risk attributable to lifestyle, we did not find higher rates of breast or colon cancer.

In particular, the increased risk of colon cancer demonstrated in this study, which was greatest for those taking antipsychotic medications, is a novel and unexpected finding. The magnitude of the risk and the degree of statistical significance does not rule out a chance finding, although it does make it unlikely. The finding of a lower risk of respiratory cancer, particularly in those who were not taking antipsychotic drugs, argues for an intrinsic protection associated with schizophrenia rather than the previously suggested effect of medication. Further research is needed to confirm or refute these findings using an independent data source with a more detailed analysis of individual drugs, including a comparison of the older class of antipsychotics with the newer atypical agents.

The higher rate of some common cancers in people with schizophrenia emphasizes the need for proactive monitoring of their physical health, which has previously been highlighted.[45] Factors such as obesity, smoking, and high alcohol consumption are more prevalent in this group[12,13] and increase the risk not only of cancer but of many other conditions, including diabetes mellitus and cardiac disease. Antipsychotic medications differ in the extent to which they increase the risk of obesity and other conditions.[46,47] If there is an inherently higher risk of some cancers in people with schizophrenia, it is particularly important to minimize any additional risks associated with lifestyle or prescribed medications.

Submitted for Publication: January 11, 2007; final revision received March 22, 2007; accepted April 27, 2007.
Correspondence: Julia Hippisley-Cox, MD, Division of Primary Care, University of Nottingham, 13th Floor Tower Bldg, University Park, Nottingham NG7 2RD, England (julia.hippisley-cox@nottingham.ac.uk).
Author Contributions: Dr Hippisley-Cox takes responsibility for the integrity of the data and the accuracy of the data analysis; all authors had full access to all the data in the study.
Financial Disclosure: None reported.
Funding/Support: This study was funded by the UK Disability Rights Commission; by core funding from the University of Nottingham (Drs Hippisley-Cox and Coupland and Ms Vinogradova); and by research and development funding from the National Health Service (Ms Parker).
Role of the Sponsor: The Disability Rights Commission had no role in the study design, data extraction, analysis, interpretation, or writing or approval of the manuscript.
Additional Contributions: We acknowledge the EMIS practices that contribute data to the QRESEARCH database free of charge. Egton Medical Information Services and David Stables, MBCHB (medical director of EMIS) provided support in creating and maintaining the research database.

REFERENCES

1. Commissioners in Lunacy for England and Wales. *Annual Report.* London, England: HMSO; 1909.
2. Barak Y, Achiron A, Mandel M, Mirecki N, Aizenberg D. Reduced cancer incidence among patients with schizophrenia. *Cancer.* 2005;104(12):2817-2821.
3. Cohen M, Dembling B, Schorling J. The association between schizophrenia and cancer: a population-based mortality study. *Schizophr Res.* 2002;57(2-3):139-146.
4. Dalton SO, Mellemkjaer L, Thomassen L, Mortensen PB, Johansen C. Risk for cancer in a cohort of patients hospitalized for schizophrenia in Denmark, 1969-1993. *Schizophr Res.* 2005;75(2-3):315-324.
5. Mortensen PB. The occurrence of cancer in first admitted schizophrenic patients. *Schizophr Res.* 1994;12(3):185-194.
6. Grinshpoon A, Barchana M, Ponizovsky A, Lipshitz I, Nahon D, Tal O, Weizman A, Levav I. Cancer in schizophrenia: is the risk higher or lower? *Schizophr Res.* 2005;73(2-3):333-341.
7. Gulbinat W, Dupont A, Jablensky A, Jensen OM, Marsella A, Nakane Y, Sartorius N. Cancer incidence of schizophrenia patients: results of record linkage studies in three countries. *Br J Psychiatry Suppl.* 1992;18(18):75-83.
8. Lichtermann D, Ekelund J, Pukkala E, Tanskanen A, Lonnqvist J. Incidence of cancer among persons with schizophrenia and their relatives. *Arch Gen Psychiatry.* 2001;58(6):573-578.
9. Brown S, Inskip H, Barraclough B. Causes of the excess mortality of schizophrenia. *Br J Psychiatry.* 2000;177:212-217.
10. Lawrence D, Holman CD, Jablensky AV, Threlfall TJ, Fuller SA. Excess cancer mortality in Western Australian psychiatric patients due to higher case fatality rates. *Acta Psychiatr Scand.* 2000;101(5):382-388.
11. Joukamaa M, Heliovaara M, Knekt P, Aromaa A, Raitasalo R, Lehtinen V. Mental

disorders and cause-specific mortality. *Br J Psychiatry*. 2001;179: 498-502.
12. McCreadie RG. Use of drugs, alcohol and tobacco by people with schizophrenia: case-control study. *Br J Psychiatry*. 2002;181:321-325.
13. McCreadie RG. Diet, smoking and cardiovascular risk in people with schizophrenia. *Br J Psychiatry*. 2003;183:534-539.
14. Jeste DV, Gladsjo J, Lindamer L, Lacro J. Medical comorbidity in schizophrenia. *Schizophr Bull*. 1996;22(3):413-430.
15. Brugha TS, Wing JK, Smith BL. Physical health of the long-term mentally ill in the community: is there unmet need? *Br J Psychiatry*. 1989;155:777-781.
16. Catts VS, Catts SV. Apoptosis and schizophrenia: is the tumour suppressor gene, p53, a candidate susceptibility gene? *Schizophr Res*. 2000;41(3):405-415.
17. Yovel G, Sirota P, Mazeh D, Shakhar G, Rosenne E, Ben-Eliyahu S. Higher natural killer cell activity in schizophrenic patients: the impact of serum factors, medication, and smoking. *Brain Behav Immun*. 2000;14(3):153-169.
18. Carrillo JA, Benitez J. Are antipsychotic drugs potentially chemopreventive agents for cancer? *Eur J Clin Pharmacol*. 1999;55(6):487-488.
19. Hippisley-Cox J, Vinogradova Y, Coupland C, Pringle M. *Comparison of Key Practice Characteristics Between General Practices in England and Wales and General Practices in the QRESEARCH Data: Report to the Health and Social Care Information Centre*. Nottingham, England: University of Nottingham; 2005.
20. Hammersley V, Hippisley-Cox J, Wilson A, Pringle M. A comparison of research general practices and their patients with other practices: cross-sectional survey in Trent. *Br J Gen Pract*. 2002;52(479):463-468.
21. Hippisley-Cox J, Hammersley V, Pringle M, Coupland C, Crown N, Wright L. How useful are General Practice databases for research? analysis of their accuracy and completeness in one research network. *Health Informatics J*. 2004; 10:91-109.
22. Goldacre MJ, Kurina LM, Wotton CJ, Yeates D, Seagroatt V. Schizophrenia and cancer: an epidemiological study. *Br J Psychiatry*. 2005;187:334-338.
23. Moore LL, Bradlee ML, Singer MR, Splansky GL, Proctor MH, Ellison RC, Kreger BE. BMI and waist circumference as predictors of lifetime colon cancer risk in Framingham Study adults. *Int J Obes Relat Metab Disord*. 2004;28(4):559-567.
24. Menvielle G, Luce D, Geoffroy-Perez B, Chastang JF, Leclerc A. Social inequalities and cancer mortality in France, 1975-1990. *Cancer Causes Control*. 2005; 16(5):501-513.
25. Fernandez E, La Vecchia C, Braga C, Talamini R, Negri E, Parazzini F, Franceschi S. Hormone replacement therapy and risk of colon and rectal cancer. *Cancer Epidemiol Biomarkers Prev*. 1998;7(4):329-333.
26. Giovannucci E. Modifiable risk factors for colon cancer. *Gastroenterol Clin North Am*. 2002;31(4):925-943.
27. Dalton SO, Johansen C, Poulsen AH, Nørgaard M, Sørensen HT, McLaughlin JK, Mortensen PB, Friis S. Cancer risk among users of neuroleptic medication: a population-based cohort study. *Br J Cancer*. 2006;95(7):934-939.
28. Halbreich U, Shen J, Panaro V. Are chronic psychiatric patients at increased risk for developing breast cancer? *Am J Psychiatry*. 1996;153(4):559-560.
29. Gulbinat W, Dupont A, Jablensky A, Jensen OM, Marsella A, Nakane Y, Sartorius N. Cancer incidence of schizophrenic patients: results of record linkage studies in three countries. *Br J Psychiatry Suppl*. 1992;18(18):75-83.
30. Oksbjerg Dalton S, Munk Laursen T, Mellemkjaer L, Johansen C, Mortensen PB. Schizophrenia and the risk for breast cancer. *Schizophr Res*. 2003;62(1-2): 89-92.
31. Halbreich U, Kinon BJ, Gilmore JA, Kahn LS. Elevated prolactin levels in patients with schizophrenia: mechanisms and related adverse effects. *Psychoneuroendocrinology*. 2003;28(suppl 1):53-67.
32. Rapp K, Schroeder J, Klenk J, Stoehr S, Ulmer H, Concin H, Diem G, Oberaigner W, Weiland SK. Obesity and incidence of cancer: a large cohort study of over 145,000 adults in Austria. *Br J Cancer*. 2005;93(9):1062-1067.
33. Kumle M, Weiderpass E, Braaten T, Persson I, Adami HO, Lund E. Use of oral contraceptives and breast cancer risk: the Norwegian-Swedish Women's Lifestyle and Health Cohort Study. *Cancer Epidemiol Biomarkers Prev*. 2002;11 (11):1375-1381.
34. Beral V; Million Women Study Collaborators. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet*. 2003;362(9382): 419-427.
35. Saku M, Tokudome S, Ikeda M, Kono S, Makimoto K, Uchimura H, Mukai A, Yoshimura T. Mortality in psychiatric patients, with a specific focus on cancer mortality associated with schizophrenia. *Int J Epidemiol*. 1995;24(2):366-372.
36. Brown LM. Epidemiology of alcohol-associated cancers. *Alcohol*. 2005;35(3): 161-168.
37. Mortensen PB. Neuroleptic medication and reduced risk of prostate cancer in schizophrenic patients. *Acta Psychiatr Scand*. 1992;85(5):390-393.
38. Jick H, Kaye J, Vasilakis-Scaramozza C, Jick SS. Risk of venous thromboembolism among users of third generation oral contraceptives compared with users of oral contraceptives with levonorgestrel before and after 1995: cohort and case-control analysis. *BMJ*. 2000;321(7270):1190-1195.
39. Hippisley-Cox J, Coupland C, Logan R. Risk of adverse gastrointestinal outcomes in patients taking cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis . *BMJ*. 2005;331(7528):1310-1316.
40. Hippisley-Cox J, Coupland C. Risk of myocardial infarction in patients taking cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis. *BMJ*. 2005;330(7504):1366.
41. Hippisley-Cox J, Coupland C. Effect of combinations of drugs on all cause mortality in patients with ischaemic heart disease: nested case control analysis. *BMJ*. 2005;330(7499):1059-1063.
42. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *BMJ*. 1991; 302(6779):766-768.
43. Nazareth I, King M, Haines A, Ranger L, Myers S. Accuracy of diagnosis of psychosis in general practice computer system. *BMJ*. 1993;307(6895):32-34.
44. Majeed A. Sources, uses, strengths and limitations of data collected in primary care in England. *Health Stat Q*. 2004;21(21):5-14.
45. Marder SR, Essock SM, Miller AL, Buchanan RW, Casey DE, Davis JM, Kane JM, Lieberman JA, Schooler NR, Covell N, Stroup S, Weissman EM, Wirshing DA, Hall CS, Pogach L, Pi-Sunyer X, Bigger JT Jr, Friedman A, Kleinberg D, Yevich SJ, Davis B, Shon S. Physical health monitoring of patients with schizophrenia. *Am J Psychiatry*. 2004;161(8):1334-1349.
46. Wirshing DA, Wirshing WC, Kysar L, Berisford MA, Goldstein D, Pashdag J, Mintz J, Marder SR. Novel antipsychotics: comparison of weight gain liabilities. *J Clin Psychiatry*. 1999;60(6):358-363.
47. Dixon L, Weiden P, Delahanty J, Goldberg R, Postrado L, Lucksted A, Lehman A. Prevalence and correlates of diabetes in national schizophrenia samples. *Schizophr Bull*. 2000;26(4):903-912.

Paper 10

Vinogradova Y, Coupland C, Hippisley-Cox J, Whyte S, Penny C. Effects of severe mental illness on survival of people with diabetes. *British Journal of Psychiatry*. 2010;197(4):272-7.

BJPsych | The British Journal of Psychiatry (2010) 197, 272–277. doi: 10.1192/bjp.bp.109.074674

Effects of severe mental illness on survival of people with diabetes

Yana Vinogradova, Carol Coupland, Julia Hippisley-Cox, Seán Whyte and Catherine Penny

Background
People with mental health problems are more likely to die prematurely than the general population but no study has examined this in individuals with diabetes.

Aims
To compare survival rates in people with diabetes with and without schizophrenia or bipolar disorder.

Method
A total of 43 992 people with diabetes were drawn from the QRESEARCH database population of over 9 million patients. Survival rates during the study period, between 1 April 2000 and 1 April 2005, and hazard ratios for deaths associated with schizophrenia and bipolar disorder were adjusted by age and gender and additionally for socioeconomic status, obesity, smoking and use of statins.

Results
Among the participants, we identified 257 people diagnosed with schizophrenia, 159 with bipolar disorder and 14 with both conditions. Although crude survival rates did not show significant differences between the groups during the study period, people with schizophrenia or bipolar disorder and diabetes, compared with those with diabetes alone, had a significantly increased risk of death after adjusting for age and gender, with hazard ratios for schizophrenia of 1.84 (95% CI 1.42–2.40) and for bipolar disorder of 1.51 (95% CI 1.10–2.07). After adjusting for the other factors, hazard ratios were 1.52 (95 CI 1.17–1.97) for schizophrenia and 1.47 (95% CI 1.07–2.02) for bipolar disorder.

Conclusions
People with schizophrenia or bipolar disorder in addition to diabetes have a relatively higher mortality rate. This suggests that diabetes either progresses more rapidly or is more poorly controlled in these individuals, or that they have higher levels of comorbidity and so are more likely to die of other causes.

Declaration of interest
None.

People with mental disorders are considerably more likely to die prematurely than the general population. In a systematic review, Harris & Barraclough[1] found that the mortality rate remained higher in individuals with mental disorders when deaths from unnatural causes (suicide and violent death) were excluded. In people with schizophrenia, the standardised mortality ratio (SMR) for all natural causes was 137% (95% CI 134–141%) compared with 100% for a general population of similar age and gender, accounting for 62% of the excess deaths from all causes. Most excess deaths were from infectious, respiratory and digestive system disorders, but deaths from endocrine, circulatory and genitourinary system disorders also had significantly raised SMRs. The SMR for endocrine system disorders, including diabetes, was 238% (95% CI 114–438%). In people with bipolar disorder, the SMR for all natural causes was 150% (95% CI 137–164%), although only deaths from circulatory and respiratory system disorders had significantly raised SMRs.

A subsequent study[2] found an SMR of 260 (95% CI 219–306) for all natural causes in people with schizophrenia, mainly as a result of diseases of the circulatory, digestive, endocrine, nervous and respiratory systems, with an SMR of 801 (95% CI 322–1651) for endocrine system disorders. Studies from two states in the USA[3,4] found the life expectancy of people with severe mental illness to be about 9 years lower than that of the general population.

The association between severe mental illness and diabetes is now widely recognised. A consensus meeting in 2003 concluded that the overall risk of type 2 diabetes in people with schizophrenia is between two and four times that of the general population, with a prevalence of approximately 15–18%, and that impaired glucose tolerance may affect up to 30% of people with schizophrenia.[5] Similar findings have been reported in people with bipolar disorder.[6]

A study in the USA examined the impact of diabetes on mortality in 197 individuals with co-occurring psychotic and substance use disorders participating in a randomised controlled study of integrated mental health and substance misuse treatment. The study found that participants with evidence of diabetes were significantly more likely to die during follow-up than participants without evidence of diabetes.[7] However, no study has yet demonstrated this in a naturalistic population sample. This study examines mortality rates in a cohort of people with diabetes, comparing those with and without schizophrenia and bipolar disorder. It tests the hypothesis that having one of these mental illnesses increases the risk of premature death in individuals with diabetes in a large and representative primary care population sample.

Method

Participants

The study used the QRESEARCH database version 8 (www.qresearch.org/), which is derived from the computerised health records of general practices using the Egton Medical Information System (EMIS) medical record computer system (Egton Medical Information Systems Limited, Leeds, UK). The full database at the time of the study contained aggregated data on more than 9 million patients from 525 representative general practices across the UK and includes information on patient demographics, diagnoses, clinical values, laboratory investigations, prescriptions, consultations and referrals. The database has been

validated by comparing birth and death rates, consultation rates and prevalence and mortality rates with other data sources,[8] and has demonstrated good levels of completeness and consistency.[9]

Inclusion criteria

QRESEARCH practices were eligible for inclusion in the study if their current EMIS computer system was installed before 1 April 1999 and they had complete data from then until 1 April 2005. Individuals from these practices were included if they: had a diagnosis of diabetes recorded between 1 January 1990 and 1 April 2000; were alive and registered with an eligible practice on 1 April 2000 (the study entry date for these analyses); had been registered with an eligible practice for at least the previous 12 months; were 25 years or older on 1 April 2000; and were not registered as a temporary resident.

Diagnoses

Individuals' diagnoses for diabetes, schizophrenia and bipolar disorder were determined using the standard computer codes (Read codes) for general practice in the UK. These codes are entered by general practitioners (GPs) when they make their clinical records of a consultation. A full list of codes, developed in consultation with GPs with an interest in the field, is available from the authors. Diagnoses were included whether they were coded as active or inactive, current or past, because the 'inactive' and 'past' codes were not reliably or consistently used by those entering data.

Type 1 and type 2 diabetes were recorded and analysed together, because the recording of diabetes in primary care does not always distinguish between types. Smoking status and obesity were defined according to the last recorded status of the individual prior to 1 April 2000. Diagnoses of schizophrenia and bipolar disorder were defined if they were recorded before 1 April 2000. Data on deaths from all causes between 1 April 2000 and 1 April 2005 were extracted from the records.

Primary outcomes and analysis

The overall survival rates for people with diabetes were determined, comparing those with and without schizophrenia and those with and without bipolar disorder. Where people had diagnoses of schizophrenia and bipolar disorder, they contributed to the analyses of both groups. The date of diagnosis of diabetes was used as the time origin in all analyses, with 1 April 2000 defined as the delayed entry date. The proportions of people alive 5 and 10 years after being diagnosed with diabetes were also determined for each population subgroup using Kaplan–Meier estimators and the differences between the groups were assessed with the log-rank test.

To estimate the risk of death in individuals with each mental disorder compared with those without, a Cox regression survival analysis was performed adjusted for age at diagnosis of diabetes (in years) and gender, and further for smoking status (smoker, non-smoker, not recorded) and body mass index ($<25\,kg/m^2$, $25–29\,kg/m^2$, $\geqslant 30\,kg/m^2$; not recorded). As it has been shown that people with schizophrenia and bipolar disorders are prescribed statins less often than the rest of the population,[10] use of statins was included. The analysis was also adjusted for Townsend deprivation score (in fifths, with higher values indicating greater deprivation), which is based on 2001 post-code-related census data, reflecting unemployment, overcrowding, lack of home ownership and lack of car ownership, and is associated with mortality in the general population.[11] In an additional analysis we also adjusted for antipsychotic use. The baseline characteristics of people with each mental disorder were compared with those without mental illness using Student's *t*-test or the chi-squared test depending on the distribution of the characteristic. All analyses were conducted in STATA version 10 for Windows.

Results

A total of 43 992 people with diabetes from 372 practices met all the inclusion criteria for the analysis. At those practices, there were a total of 1 896 944 patients aged 25 or over on the study entry date who had been permanently registered for at least 12 months, giving a prevalence of diabetes in those aged 25 and over of 2.3%. Of the 43 992 people with diabetes, 257 (0.58%) had a diagnosis of schizophrenia, 159 (0.36%) had a diagnosis of bipolar disorder and 14 were diagnosed with both conditions (these individuals are also included in the 257 and 159 figures above). These proportions are broadly comparable with other UK population surveys.[12,13]

Table 1 and Fig. 1 show the characteristics of the individuals with schizophrenia, bipolar disorder and with neither condition. The proportion of women was higher in both mental illness groups: 55% of individuals with schizophrenia were female and 60% of those with bipolar disorder were female compared with 46% of those with diabetes alone. Participants in both mental illness groups were less likely to live in rural areas, more likely to live in deprived areas (Townsend score fifths 4 or 5) and to smoke, and – for individuals with schizophrenia – to have a higher body mass index. Individuals in both mental illness groups had a lower mean systolic blood pressure.

People with schizophrenia were younger at study entry and when diagnosed with diabetes. In total, 48% were aged 60 years or older at diagnosis compared with 59% of those without schizophrenia or bipolar disorder. Use of statins in the 12 months prior to the study reference date was lower in individuals with both schizophrenia and diabetes than in the population with diabetes alone (10% *v.* 17%).

Survival rates

During the study period, 8698 people died. Of these, 57 had schizophrenia, 39 had bipolar disorder and 1 had both conditions. In total, 22% of those with schizophrenia and 25% of those with bipolar disorder died during follow-up compared with 20% with diabetes alone.

Table 2 shows 5- and 10-year survival rates for individuals with diabetes and schizophrenia or bipolar disorder and with diabetes alone; these are estimates of the proportions of individuals who are still alive 5 years and 10 years after being diagnosed with diabetes. Although people with these mental health problems had lower survival rates than individuals with diabetes alone, the differences in the survival rates were not statistically significant for either schizophrenia or bipolar disorder.

Hazard ratios

Table 3 presents the hazard ratios for dying in those with diabetes, comparing individuals with and without schizophrenia and bipolar disorder. The adjusted hazard ratios show the risks taking into account the differences in age and gender and also the differences in deprivation, obesity, smoking habits and use of statins between the groups. Whereas the unadjusted hazard ratios do not reach statistical significance, the adjusted hazard ratios are significantly increased for people with schizophrenia and those with bipolar disorder compared with those with diabetes alone.

Table 1 Characteristics of individuals with diabetes: numbers (%) and means (s.d.) in the groups with schizophrenia and bipolar disorder and without these mental illnesses

	With schizophrenia ($n = 257$)	With bipolar disorder ($n = 159$)	No schizophrenia or bipolar disorder ($n = 43\,589$)
Female, *n* (%)	141 (54.9)*	96 (60.4)*	19 854 (45.5)
Male, *n* (%)	116 (45.1)*	63 (39.6)*	23 736 (54.5)
Age on 1 April 2000, years: mean (s.d.)	60.9 (12.5)*	64.5 (13.0)	65.2 (13.8)
Body mass index, kg/m^2: mean (s.d.)	29.8 (6.4)*	30.0 (6.2)	29.0 (5.8)
Systolic blood pressure, mmHg: mean (s.d.)	138.4 (21.0)*	139.2 (19.3)*	144.8 (20.1)
Diastolic blood pressure, mmHg: mean (s.d.)	80.9 (11.3)	80.5 (9.1)	81.2 (10.1)
Age at diagnosis of diabetes	*		
Age, years: mean (s.d.)	57.2 (12.6)*	60.4 (13.1)	61.1 (13.9)
Under 50 years, *n* (%)	67 (26.1)	30 (18.9)	8563 (19.6)
50–59 years, *n* (%)	68 (26.5)	34 (21.4)	9566 (21.9)
60–69 years, *n* (%)	82 (31.9)	58 (36.5)	12 666 (29.1)
70 years plus, *n* (%)	40 (15.6)	37 (23.3)	12 795 (29.4)
Townsend score: fifths, *n* (%)	*	*	
1 (least deprived)	19 (7.4)	20 (12.6)	7927 (18.2)
2	26 (10.1)	21 (13.2)	7853 (18.0)
3	34 (13.2)	22 (13.8)	8342 (19.1)
4	49 (19.1)	40 (25.2)	8477 (19.4)
5 (most deprived)	120 (46.7)	45 (28.3)	8969 (20.6)
Lives in rural area	52 (20.2)*	37 (23.3)*	15 165 (34.8)
Smoking status, *n* (%)	*	*	
Smoker	84 (32.7)	46 (28.9)	7054 (16.2)
Non-smoker	132 (51.4)	103 (64.8)	32 739 (75.1)
Not recorded	41 (16.0)	10 (6.3)	3796 (8.7)
Medication, *n* (%)			
On antipsychotics prior to study period	168 (65.4)*	68 (42.8)*	4925 (11.3)
On statins prior to study period	26 (10.1)*	24 (15.1)	7561 (17.3)
Died during study period, *n* (%)	57 (22.2)	39 (24.5)	8603 (19.7)

*Significantly different from the group without mental illness, $P<0.01$.

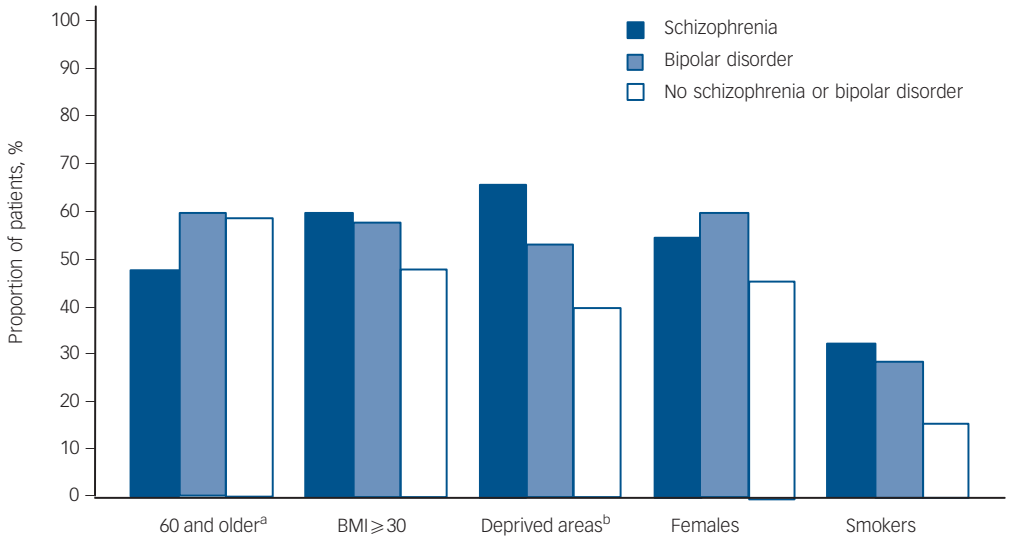

Fig. 1 Proportion of individuals with basic characteristics.

a. Age 60 or older at diagnosis of diabetes.
b. Deprived areas with Townsend score fifths of 4 and 5.

The adjusted hazard ratios were 1.52 (95 CI 1.17–1.97) for schizophrenia and 1.47 (95% CI 1.07– 2.02) for bipolar disorder. After additional adjustment for overall use of antipsychotics, the effects of schizophrenia and bipolar disorders on mortality were diminished but remained statistically significant (adjusted odds ratios 1.38, 95% CI 1.06–1.80 and 1.41, 95%CI 1.03–1.94 respectively).

Figure 2 shows the estimated survival proportions of over a number of years based on a diagnosis of diabetes at 60 years of age. The estimates were based on the model where only gender

Table 2 Crude survival rates (95% CI) after diagnosis of diabetes

	% (95% CI)		
	With schizophrenia ($n = 257$)	With bipolar disorder ($n = 159$)	No schizophrenia or bipolar disorder ($n = 43\,589$)
Crude survival rates at 5 years after diabetes diagnosis	79 (69–87)	91 (80–96)	82 (81–83)
Crude survival rates at 10 years after diabetes diagnosis	63 (53–71)	60 (49–70)	65 (64–67)

Table 3 Hazard ratios (95% CI) for mortality during follow-up for people with schizophrenia and bipolar disorder between 1 January 2000 and 1 April 2005

	Hazard ratios (95% CI)		
	Unadjusted	Adjusted for age[a] and gender	Adjusted for all factors[b]
Individuals without schizophrenia or bipolar disorder	1.00	1.00	1.00
Individuals with schizophrenia	1.21 (0.93–1.57)	1.84 (1.42–2.40)	1.52 (1.17–1.97)
Individuals with bipolar disorder	1.30 (0.95–1.78)	1.51 (1.10–2.07)	1.47 (1.07–2.02)

a. Age at diagnosis of diabetes.
b. Adjusted for age at diagnosis of diabetes, gender, smoking status, deprivation, obesity, use of statins.

and age were taken into account (Table 3). Survival rates were lowest for the people with schizophrenia, then for the group with bipolar disorder and were highest for those without these mental health problems.

Discussion

Findings

In line with previous research,[14–16] these results show that people with diabetes who also had schizophrenia or bipolar disorder, were more likely to live in more deprived areas, to be smokers and to be female than those who did not have these disorders. Individuals with schizophrenia or bipolar disorder had a lower mean systolic blood pressure, which could be explained by the younger age and higher proportion of women.[17] People with schizophrenia were also younger at diagnosis of diabetes and had a higher body mass index.

An analysis of Kaplan–Meier estimates of survival rates showed no significant differences between those with and without the mental disorders. However, people with schizophrenia were younger and those in both groups were more likely to be female so, after adjusting for age and gender, individuals with diabetes and schizophrenia or bipolar disorder had a significantly increased risk of death compared with those with diabetes alone. Adjusting for additional confounding variables (smoking status, deprivation, obesity, use of statins) slightly reduced the risk of death in people with schizophrenia and diabetes, but it still remained 50% higher (the adjusted hazard ratio was 1.5) than for those with diabetes alone. The risk of death for people with bipolar disorder adjusted for age and gender was 50% higher than for those without these mental health problems and changed very little after adjusting for the other confounders.

We have shown that people with diabetes and schizophrenia or bipolar disorder have higher mortality rates than individuals with diabetes alone. There are several possible explanations for this difference. In people with schizophrenia or bipolar disorder, diabetes may progress more rapidly, it may be more poorly controlled, or these individuals may have more comorbid physical illnesses and therefore be more likely to die of other causes.

In the UK, it seems unlikely that people with schizophrenia or bipolar disorder now have more poorly controlled diabetes. A

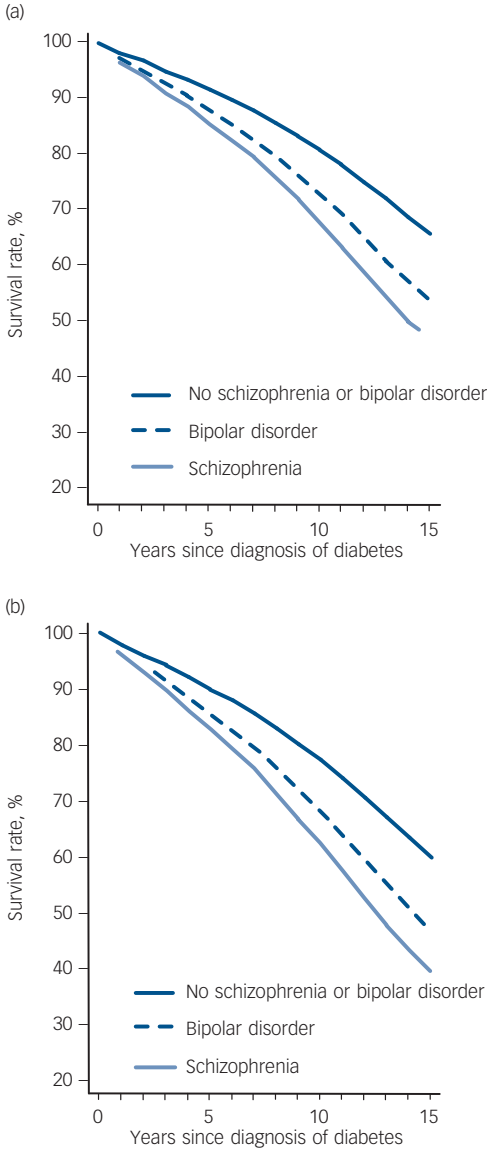

Fig. 2 Estimated survival rates in (a) women and (b) men with diabetes at 60 years of age.

recent study was conducted as part of the same Disability Rights Commission project[16] as this study and used a similar sample derived from the same database. It found that, since the introduction of the new General Medical Services contract for general practitioners on 1 April 2004, diabetes care in the UK has been as good for people with schizophrenia or bipolar disorder as for those without and, in particular, that they had as good or better glucose control as measured by HbA_{1c}. The new contract was, however, introduced towards the end of the period from which data for this study was collected so it is possible that prior to its introduction diabetes care was worse for those with schizophrenia or bipolar disorder, partly or wholly accounting for the difference in mortality seen in this study.

It seems more likely that the increased mortality rate is related to greater rates of comorbid physical illnesses in individuals with schizophrenia and bipolar disorder. This has been found in other studies[1,18] and, in relation to chronic obstructive pulmonary disease, stroke, coronary heart disease and certain cancers, in some studies associated with the Disability Rights Commission project.[10,19]

Limitations

This study only included people who were registered with a GP, so it cannot comment on people who do not access primary care and who may have different rates of diabetes, schizophrenia or bipolar disorder. These include some prisoners and people with schizophrenia or bipolar disorder known to a community mental health team but not to primary care services. It also cannot comment on people who have diabetes, schizophrenia or bipolar disorder, but who have not been diagnosed as such. There are, for example, an estimated 600 000 people in the UK with undiagnosed diabetes.[13]

A greater concern is that because people with schizophrenia and bipolar disorder have a shorter life expectancy, on average more of them might have died before being included in the study than might those with diabetes alone. It is, therefore, possible that the survivors are not typical of the whole population with schizophrenia or bipolar disorder.

The survival analyses were based on all-cause mortality rather than deaths attributed to diabetes. This avoids inaccuracy as a result of the known unreliability of recorded causes of death, but prevents differentiation between a diabetes-related cause of increased mortality and increased mortality due to comorbidity. A high proportion of the people with mental disorders in this study were prescribed antipsychotic drugs and it is possible that these might have increased or decreased their survival rates depending on the type of drug and its duration of use, as shown in a study of mortality in people with schizophrenia.[20] However, we did not carry out a detailed analysis according to type of antipsychotic drug prescribed, dose or duration of use as this was not within the scope of this study, but, adjusting for the overall use of antipsychotics in the survival analysis did not explain the increased mortality in the groups with bipolar disorder or schizophrenia.

These findings demonstrate the importance of good-quality diabetes care for people with schizophrenia or bipolar disorder. Individuals with schizophrenia or bipolar disorder have both an increased prevalence of diabetes and lower survival rates after diabetes is diagnosed. Diagnosing and treating people with schizophrenia or bipolar disorder proactively and ensuring that they take advantage of the healthcare available to them is necessary to reduce the inequality in outcomes between those with and without these mental illnesses.

Yana Vinogradova, MSc, **Carol Coupland**, PhD, **Julia Hippisley-Cox**, PhD, MD, Division of Primary Care, University of Nottingham, University Park, Nottingham; **Seán Whyte**, PhD, Springfield Hospital, London; **Catherine Penny**, PhD, Institute of Psychiatry, London, UK

Correspondence: Yana Vinogradova, Department of Primary Care, Tower Building 13th Floor, University Park, University of Nottingham, Nottingham NG7 2RD, UK. Email: yana.vinogradova@nottingham.ac.uk

First received 23 Oct 2009, final revision 23 Apr 2010, accepted 14 Jun 2010

Funding

This study was funded by the UK Disability Rights Commission and by core funding from the University of Nottingham.

Acknowledgements

We thank the practices contributing to the QRESEARCH database and also acknowledge the technical expertise of David Stables (EMIS computing) in establishing the QRESEARCH database.

References

1 Harris EC, Barraclough B. Excess mortality of mental disorder. *Br J Psychiatry* 1998; **173**: 11–53.

2 Brown S, Kim M, Mitchell C, Inskip H. Twenty-five year mortality of a community cohort with schizophrenia. *Br J Psychiatry* 2010; **196**: 116–21.

3 Dembling BP, Chen DT, Vachon L. Life expectancy and causes of death in a population treated for serious mental illness. *Psychiatr Serv* 1999; **50**: 1036–42.

4 Tsuang MT, Woolson RF, Fleming JA. Premature deaths in schizophrenia and affective disorders. *Arch Gen Psychiatry* 1980; **37**: 979–83.

5 Expert Group. 'Schizophrenia and diabetes 2003'. Expert Consensus Meeting, Dublin 3–4 October 2003: consensus summary. *Br J Psychiatry* 2004; **184** (suppl 47): s112–4.

6 Morriss R, Mohammed FA. Metabolism, lifestyle and bipolar affective disorder. *J Psychopharmacol* 2005; **19** (suppl 6): 94–101.

7 Jackson C, Covell N, Drake R, Essock S. Relationship between diabetes and mortality among persons with co-occurring psychotic and substance use disorders. *Psychiatr Serv* 2007; **58**: 270–2.

8 Hammersley V, Hippisley-Cox J, Wilson A, Pringle M. A comparison of research General Practices and their patients with other practices – cross sectional survey in Trent. *Br J Gen Pract* 2002; **52**: 463–8.

9 Hippisley-Cox J, Hammersley V, Pringle M, Coupland C, Crown N, Wright L. How useful are General Practice databases for research? Analysis of their accuracy and completeness in one research network. *Health Informatics J* 2004; **10**: 91–109.

10 Hippisley-Cox J, Parker C, Coupland C, Vinogradova Y. Inequalities in the primary care of patients with coronary heart disease and serious mental health problems: a cross-sectional study. *Heart* 2007; **93**: 1256–62.

11 Office for National Statistics. *Geographic Variations in Health*: 326–37. TSO (The Stationery Office), 2001.

12 Jenkins R, Bebbington P, Brugha TS, Farrell M, Lewis G, Meltzer H. British psychiatric morbidity survey. *Br J Psychiatry* 1998; **173**: 4–7.

13 The Information Centre for Health and Social Care. *Key Findings about the Quality of Care for People with Diabetes in England, incorporating Registrations from Wales. National Diabetes Audit*. The Information Centre, National Clinical Auidt Support Programme, 2006.

14 Dickerson FB, Brown CH, Kreyenbuhl JA, Fang L, Goldberg RW, Wohlheiter K, et al. Obesity among individuals with serious mental illness. *Acta Psychiatr Scand* 2006; **113**: 306–13.

15 Kelly C, McCreadie RG. Smoking habits, current symptoms, and premorbid characteristics of schizophrenic patients in Nithsdale, Scotland. *Am J Psychiatry* 1999; **156**: 1751–7.

16 Whyte S, Penny C, Phelan M, Hippisley-Cox J, Majeed A. Quality of diabetes care in patients with schizophrenia and bipolar disorder: cross-sectional study. *Diabet Med* 2007; **24**: 1442–8.

17 The NHS Information Centre for Health and Social Care. *Health Survey for England – 2008 Trend Tables*. Joint Health Surveys Unit, 2008.

18 Phelan M, Stradins L, Morrison S. Physical health of people with severe mental illness. *BMJ* 2001; **322**: 443–4.

19 Hippisley-Cox J, Vinogradova Y, Coupland C, Parker C. Risk of malignancy in patients with schizophrenia or bipolar disorder: nested case-control study. *Arch Gen Psychiatry* 2007; **64**: 1368–76.

20 Tiihonen J, Lönnqvist J, Wahlbeck K, Klaukka T, Niskanen L, Tanskanen A, et al. 11-year follow-up of mortality in patients with schizophrenia: a population-based cohort study (FIN11 study). *Lancet* 2009; **374**: 620–7.

Paper 11

Vinogradova Y, Hippisley-Cox J, Coupland C. Identification of new risk factors for pneumonia: population based case control study. *British Journal of General Practice*. 2009;59(567):742-9.

Paper 12

Vinogradova Y, Coupland C, Hippisley-Cox J. Risk of pneumonia in patients taking statins: population-based nested case-control study.
British Journal of General Practice. 2011;61(592):e742-e8(7).

Statement about joint authorship

Paper 1

Vinogradova Y, Hippisley-Cox J, Coupland C, Logan R. Risk of colorectal cancer in patients prescribed statins, nonsteroidal anti-inflammatory drugs, and cyclooxygenase-2 Inhibitors: nested case-control study. *Gastroenterology*. 2007;133:393–402.

This publication describes research into associations between the use of some common drugs and colorectal cancer. I contributed to the detailed analysis plan, in particular with respect to defining exposure, and I performed all of the statistical analysis, interpreting the results and contributing to the final document by writing the first drafts of Method, Results and Discussion with input from my co-authors.

Paper 2

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to statins and risk of common cancers: a series of nested case-control studies. *BMC Cancer*. 2011;11:409.

This paper describes an investigation into associations between statin use and the risks of the most common cancers. I did the literature review, developed the study design and the detailed analysis plan, performed all the data manipulation and statistical analysis with input from my co-authors. I took the lead in drafting the paper and in revising the manuscript and responding to reviewers in the publication process.

Paper 3

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to cyclo-oxygenase-2 inhibitors and risk of cancer: nested case-control studies. *British Journal of Cancer*. 2011;105:452-9.

This describes a study looking into associations between use of COX2 inhibitors and risks of the most common cancers. I did the literature review, developed the detailed analysis plan, performed the data manipulation and statistical analysis, interpreted the results and wrote the paper, with input from my co-authors. In the publication peer-review process, I led revision of the manuscript and responses to reviewers.

Pa*p*er 4

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of cancer: a protocol for nested case–control studies using the QResearch primary care database. *BMJ Open*. 2012, 2012; 2(1):e000548.

This is a protocol for a proposed study of associations between bisphosphonate use and the risks of the most common cancers and an additional three female-specific cancers. I undertook literature review, developed the detailed analysis plan, providing justification for the inclusion of various factors into the models for different cancers, performed the analysis, and wrote the manuscript and took a lead in the publication process with input from my co-authors.

Papers 5 and 6

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of gastrointestinal cancers: series of nested case-control studies with QResearch and CPRD data. BMJ. 2013;346:f114.

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of non-gastrointestinal cancers: series of nested case-control studies using QResearch and CPRD data. *British Journal of Cancer*. 2013;109:795-806.

These papers are two publications arising from the project proposed in the Paper 4 protocol, where the results for gastrointestinal and for non-gastrointestinal cancers were reported separately. I was responsible for the data extraction from CPRD and, using the data from QResearch and CPRD, I performed all the data manipulation, ran the statistical analyses according to the published protocol, and drafted and developed the papers with input from my co-authors. I also managed the publication peer review process.

Paper 7

Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to combined oral contraceptives and risk of venous thromboembolism: a protocol for nested case–control studies using the QResearch and the CPRD databases. *BMJ Open* 2014;4(4):e004499

This is a protocol for a study of the risks of venous thromboembolism associated with use of combined oral contraceptives. I did the literature review, developed the study design and wrote the manuscript with input from my co-authors, justifying the need for a new study in a previously researched area and again proposing the use of data from QResearch and CPRD to improve risk estimates. I also took the lead in responding to peer reviewers.

Paper 8

Vinogradova Y, Coupland C, Hippisley-Cox J. Combined oral contraceptives and risk of venous thromboembolism: nested case–control studies using the QResearch and the CPRD databases. *BMJ* 2015 350:h2135

This paper is the publication arising from the research proposed in Paper 7 into the associations between venous thromboembolism and use of combined oral contraceptives in younger women. As in the bisphosphonates study (Papers 5 and 6), I managed some of the required data extraction – in this case from CPRD and HES. I undertook all subsequent data manipulation and statistical analyses, interpreted the results, and developed the paper with input from my co-authors. I also took the lead in a fairly challenging peer review process.

Paper 9

Hippisley-Cox J, **Vinogradova Y**, Coupland C, Parker C. Risk of Malignancy in Patients With Schizophrenia or Bipolar Disorder: Nested Case-Control Study. *Archives of General Psychiatry*. 2007, 2007;64(12):1368-76.

This paper describes is a study investigating risks of five common cancers in patients with severe mental illness. Here I performed the statistical analysis, presented the results and contributed to writing the paper (writing Methods and Results).

Paper 10

Vinogradova Y, Coupland C, Hippisley-Cox J, Whyte S, Penny C. Effects of severe mental illness on survival of people with diabetes. *British Journal of Psychiatry*. 2010;197(4):272-7

This publication describes research into the survival of diabetes patients, with and without serious mental illness. I contributed to the design of the study, performed the statistical analysis, interpreted the results and wrote Methods and Results sections. I also reviewed and commented on manuscript drafts.

Paper 11

Vinogradova Y, Hippisley-Cox J, Coupland C. Identification of new risk factors for pneumonia: population based case control study. *British Journal of General Practice*. 2009;59(567):742-9.

This paper presents results from a study identifying morbidities associated with increased risk of pneumonia. In this, I contributed to the study design, performed the data manipulation and statistical analysis, interpreted the results and developed the paper with input from my co-authors.

Paper 12

Vinogradova Y, Coupland C, Hippisley-Cox J. Risk of pneumonia in patients taking statins: population-based nested case-control study.
British Journal of General Practice. 2011;61(592):e742-e8(7).

This paper describes an investigation into the associations between use of statins and risk of pneumonia in older patients. For this study, I undertook the literature review, prepared the detailed analysis plan, performed all of the data manipulation and statistical analysis, and wrote the manuscript, with input from my co-authors.