A combined connectivity mapping and pharmacoepidemiology approach to identify existing medications with breast cancer causing or preventing properties

Busby, J., Murray, L., Mills, K., Zhang, S.-D., Liberante, F., & Cardwell, C. (2018). A combined connectivity mapping and pharmacoepidemiology approach to identify existing medications with breast cancer causing or preventing properties. *Pharmacoepidemiology and Drug Safety*, *27*(1), 78-86. https://doi.org/10.1002/pds.4345

Published in:
Pharmacoepidemiology and Drug Safety

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Title: A combined connectivity mapping and pharmacoepidemiology approach to identify existing medications with breast cancer causing or preventing properties

Running title: Connectivity mapping for medication repurposing

Authors: John Busby[1], Liam Murray[1], Ken Mills[2], Shu-Dong Zhang[2,3], Fabio Liberante[2], Chris Cardwell[1]

[1] Centre for Public Health, Queen's University Belfast, Belfast, UK

[2] Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, UK

[3] Northern Ireland Centre for Stratified Medicine, Biomedical Sciences Research Institute, University of Ulster, C-TRIC Building, Altnagelvin Area Hospital, Londonderry, UK

Correspondence: Dr. John Busby, Centre for Public Health, Queen's University Belfast, Institute for Clinical Science, Royal Victoria Hospital, Grosvenor Road, Belfast, BT12 6BJ, +442890971640, john.busby@qub.ac.uk

Keywords: drug discovery, connectivity mapping, pharmacoepidemiology, breast cancer

Key points:

- Several commonly used medications have been previously found to alter cancer risk. However identifying these is difficult.
- We applied a novel combined connectivity mapping and pharmacoepidemiological approach to identify new medications which alter breast cancer risk.
- Overall, our combined connectivity mapping and pharmacoepidemiological approach did not identify any additional medications which were substantially associated with breast cancer risk.
- Additional work exploring the causes of our null results are required to refine this methodology for future studies.

Manuscript word count: 3,051

Abstract word count: 219

Abstract

Purpose: We applied a novel combined connectivity mapping and pharmacoepidemiological approach to identify medications which alter breast cancer risk.

Methods: The connectivity mapping process identified six potentially cancer-causing (meloxicam, azithromycin, rizatriptan, citalopram, rosiglitazone, verapamil) and four potentially cancer-preventing (bendroflumethiazide, sertraline, fluvastatin, budesonide) medications which were suitable for pharmacoepidemiological investigation. Within the UK Clinical Practice Research Datalink, we matched 45,147 breast cancer cases to one control per case based on age, year and GP practice. Medication use was determined from electronic prescribing records. We used conditional logistic regression to calculate odds ratios (ORs) for the association between medication use and cancer risk after adjustment for comorbidities, lifestyle factors, deprivation and other medication use.

Results: Bendroflumethiazide was associated with increased breast cancer risk (OR: 1.11; 95% CI: 1.06, 1.15) however the connectivity mapping exercise predicted this medication would reduce risk. There were no statistically significant associations for any of the other candidate medications, with ever use ORs ranging from 0.93 (95% CI: 0.78, 1.11) for azithromycin to 1.16 (95% CI: 0.0.99, 1.37) for verapamil.

Conclusions: In this instance, our combined connectivity mapping and pharmacoepidemiological approach did not identify any additional medications which were substantially associated with breast cancer risk. This could be due to limitations in the connectivity mapping, such as implausible dosage requirements, or the pharmacoepidemiology, such as residual confounding.

Introduction

Breast cancer is the second most common cancer in the world, with 1.7 million new cases diagnosed each year.[1, 2] Survival rates are 85% at five years[3], with patients suffering reduced quality of life during treatment and recovery.[4-6] The disease places a large financial burden on society; in Europe it accounts for €6.7 billion in direct healthcare costs, and a further €8.2 billion in lost productivity and informal care costs.[7] Additional prevention and treatment strategies are required to improve patient outcomes and healthcare finances.

Recently, there has been much interest in exploring new uses for existing medications [8, 9], with notable successes in cancer treatment. For instance, aspirin has been shown to prevent colorectal cancer in certain high risk patients.[10] Identifying medications with previously unrecognised cancer-causing properties could also improve prevention, particularly when prescribing is discretionary. For example, reports of higher breast cancer risk among hormone replacement therapy users have led to a sharp decline in its use.[11, 12] Despite the benefits of identifying existing medications which alter breast cancer risk, the large number of potential candidates for evaluation makes prioritising investigations difficult.

Connectivity mapping is an advanced bioinformatics technique that can be used to identify medications which mimic or reverse the gene-expression profiles induced by a disease.[13, 14] It has previously been used to identify medications with anti-cancer properties. For instance, cimetidine has been discovered as a potential treatment for lung cancer[15], and rapamycin has been shown to overcome dexamethasone resistance in acute lymphoblastic leukaemia.[14] Pharmacoepidemiology allows the study of the association between medication use and outcomes in humans, quickly and at modest cost.[16] Therefore, a combination of connectivity mapping and pharmacoepidemiology could be used to rapidly identify medications which could alter breast cancer risk. This paper aims to apply this approach using population-based data from the UK.

Methods

Initial connectivity mapping screening and results

Our application of the connectivity mapping is described briefly in Appendix 1, and more fully elsewhere.[17] Further details of this process are available upon request. The primary outcome of this was the 'connectivity score'[14], which is similar to a z-score from a standard normal distribution, and can be interpreted as the strength of agreement between the gene signature of the disease and medication. In this instance, a positive connectivity score indicates a medication with potential cancer-causing properties, whilst a negative score signifies a medication with potential cancer-preventing properties.

We initially identified 67 medications for inclusion in the analysis[17], but excluded 57 of these (Appendix 2) as they were rarely prescribed within primary care (n=28, 42%), already used in cancer treatment (n=21, 31%), typically given in topical or nasal spray formulations which are unlikely to have a systemic effect (n=5, 7%), widely available over-the-counter (n=2, 3%), or already known to influence breast cancer risk (n=1, 1%). Of the ten remaining substances, six were identified as potentially increasing (meloxicam, azithromycin, rizatriptan, citalopram, rosiglitazone, verapamil) and four as potentially reducing (bendroflumethiazide, sertraline, fluvastatin, budesonide) the risk of breast cancer.

Pharmacoepidemiology data

We conducted a nested case-control study using the Clinical Practice Research Datalink (CPRD). The CPRD contains computerised medical records from 674 general practices (approximately 7% of the UK population).[18] Practices are audited by the CPRD and those meeting a predefined standard on data completeness and quality are deemed 'up to research standard', and included in future extracts. Data recorded includes patient demographics, clinical diagnoses (using Read codes) and prescription medication use. Previous research has found CPRD prescription and clinical information to be of high quality.[18-20] CPRD data was linked to the National Cancer Data Repository (NCDR)[21] to identify patients with a registry-confirmed breast cancer diagnosis, and to census information to derive Index of Multiple Deprivation (IMD) scores. The NCDR holds UK-wide data on cancer registration from a variety of sources including general practices, cancer screening programmes, NHS and private hospitals, and death certificates.[21]

Ethical approval for all purely observational research using anonymised CPRD data was obtained from a National Research Ethics Service Committee (NRECS). The protocol for this study was approved by the CPRD Independent Scientific Advisory Committee (Ref: 15_212R), and has been made available to reviewers.

Cases and controls

In our primary analysis, cases were defined as patients with a first ever breast cancer diagnosis (ICD 10 code C50.0 to C50.9), identified from the NCDR cancer registry or GP records, between 1st January 1995 and 31st December 2010. This reflected a slight change from our initial analysis plan, which restricted cases to those with an NCDR-confirmed breast cancer diagnosis. We decided to use a broader case definition as recent evidence suggests that over 96% of breast cancers recorded within the CPRD can be validated using other data sources.[22] We conducted a secondary analysis restricted to NCDR-confirmed cases. Each case was matched to one control, who was breast cancer free on the cancer diagnosis date of the case, based on age (+/- one year if no exact match available), year of diagnosis and GP practice. The index date for the cases was their breast cancer diagnosis date. The index date of the controls was set equal to their matched case. The start of the exposure period was the latest of the patient's registration date at the practice, or when the practice's records were deemed to be 'up to research standard'. Cases and controls were included in the study if they had at least three years of 'up to research standard' medical records before their index date. This was to ensure that there was adequate time to ascertain exposure status and measure confounders. Where cases and controls had different lengths of follow-up, these were truncated to the shortest period within the matched set to avoid time-window bias.[23]

Definition of exposure

We used the British National Formulary (August 2016 version) to compile a list proprietary and generic drug names containing the compounds identified from the connectivity mapping (Appendix 3). Prescriptions in the year prior to the index date were excluded to prevent reverse causation.[24] We defined patients as users if they had at least one prescription for the candidate medication during the exposure period. To enable the testing of dose-response relationships we extracted data on the medication prescribed, number of packs / tablets and medication strength, and calculated defined daily doses (DDDs). The DDD system is a validated measure of drug consumption maintained by the World Health Organisation.[25] A single DDD is the average maintenance dose per day of a drug used for its main indication in adults (e.g. depression for citalopram). There was insufficient information to calculate DDDs for 0.4% of prescriptions, and implausible values were recorded in a further 2.0% (e.g. 1 tablet; >50,000 tablets). In these cases we assumed the most common DDD

based on other prescriptions of that medication with complete information. We calculated the total number of DDDs received during follow-up, and categorised patients into those receiving zero, between 1 and 365 DDDs (less than one year's usage), and greater than 365 DDDs (more than one year's usage).

Accounting for confounding factors

We used data recorded in the CPRD during the exposure period to adjust for body mass index (normal [BMI under 25], overweight [BMI between 25 and 30], obese [BMI over 30]), smoking status (none, ex, current) and alcohol use (none, ex, current). We adjusted for differences in deprivation using the IMD score of the patient's home address. We accounted for the presence of ductal carcinoma in situ (DCIS; Read codes: B830.00, B825000, B830100) and each of the fifteen non-cancer related comorbidities within the Charlson index[26] during the exposure period (AIDS, cerebrovascular disease, chronic pulmonary disease, congestive heart disease, dementia, diabetes, diabetes with complications, hemiplegia, mild liver disease, moderate liver disease, myocardial infarction, peptic ulcer disease, peripheral vascular disease, renal disease, rheumatological disease) using a list of Read codes modified for use in the CPRD.[27] We also adjusted for the use of aspirin, statins, metformin, digoxin, oral contraceptives and hormone replacement therapy, as associations with breast cancer risk have been identified previously (Appendix 4).[8, 28, 29] We ignored confounders in the year prior to diagnoses to reduce reverse causation and overadjustment bias.[30]

Statistical Analysis

We calculated descriptive statistics and compared the demographics and clinical characteristics of the cases and controls. We used conditional logistic regression to estimate odds ratios (ORs) with 95% confidence intervals (CIs) for the association between medication use and breast cancer. We ranked medications by their connectivity score and plotted this against their odds ratio to summarise our results. We used Spearman rank correlation coefficient to formally test for an association between the connectivity score and effect estimate. The matched design of the study appropriately accounted for the effect of age, general practice and year of diagnosis. We made additional adjustments for the presence of each comorbidity (yes/no), DCIS (yes/no), lifestyle factors (categorical), deprivation (quintiles) and other medication use (yes/no) using regression.

Although our main analysis was complete case, we also performed an additional sensitivity analysis using multiple imputation with chained equations (MICE) for smoking, alcohol consumption and BMI. The imputation used ordered logit models with age and deprivation, separately for cases and controls. Briefly, MICE is a simulation-based approach for handling missing data which leads to valid

statistical inferences.[31] Sensitivity analyses were also conducted investigating the impact of excluding prescriptions in the two years prior to the index date (as opposed to one in the main analysis) and defining medication users as patients with at least three prescriptions (as opposed to one in the main analysis). Finally, we used Wald interaction tests to compare associations by age (dichotomised at 50 as this corresponds to the median menopausal age within the UK [32, 33]), and estrogen receptor status (based on receipt of tamoxifen or aromatase inhibitors in the six months after breast cancer diagnosis), as risk factors have been shown to differ between these groups previously.[34, 35]

Results

Patient characteristics

We identified 45,147 cases of breast cancer in our primary analysis with a median exposure period of 7.8 years (min: 3.0, max: 27.1). Overall, cases and controls were very similar in terms of their demographics, lifestyle and clinical history, although a slightly larger proportion of cases had previously been prescribed hormone replacement therapy (28.5% vs. 26.2%) and oral contraceptives (7.5% vs. 6.6%) (Table 1).

Association between medication use and breast cancer

Overall, there was little convincing evidence that any of the medications included in our study substantially altered breast cancer risk (Table 2). The adjusted odds ratios (ORs) for ever use of azithromycin, rizatriptan, citalopram, rosiglitazone, sertraline, fluvastatin and budesonide were between 0.95 and 1.05. There was some evidence that verapamil (OR: 1.16; 95% CI: 0.99, 1.37) and bendroflumethiazide use (OR: 1.11; 95% CI: 1.06, 1.15) was associated with increased breast cancer risk, although effect estimates were small. For both these medications, our results were inconsistent with a dose-response relationship. The odds ratios for each of the confounders included in our model are provided in Appendix 5.

There was no clear relationship between the connectivity mapping score and the adjusted ORs for breast cancer risk (Spearman's correlation coefficient= -0.067; p-value=0.855). The five medications with the largest connectivity scores (meloxicam, azithromycin, rizatriptan, citalopram and rosiglitazone), and three with the lowest connectivity scores (budesonide, fluvastatin and sertraline), had a small association with breast cancer, even among patients who received more than one year of treatment (Table 2, Figure 1). Furthermore, bendroflumethiazide was identified as a potentially cancer-reducing medication in the connectivity mapping, yet we found an 11% (OR: 1.11; 95% CI: 1.06, 1.15) increase in breast cancer risk amongst users. Our findings were similar when restricting cases to those with a NCDR-confirmed breast cancer diagnosis, although the dose-response relationship for verapamil was no longer apparent. Furthermore, due to the smaller sample size, confidence intervals were noticeably wider in this analysis (Table 2, Appendix 6).

Sensitivity and subgroup analysis

In general, our conclusions were little altered in sensitivity analyses (Table 3). Similar associations were observed when prescriptions were excluded in the two years prior to diagnosis (rather than one), and when the definition of medication 'ever use' was based upon three or more prescriptions

(rather than one). Using MICE to adjust for lifestyle factors (smoking, alcohol consumption, and obesity) resulted in similar estimates to the main, complete-case analysis. There was little evidence that any of the associations observed in our study differed according to menopausal or estrogen-receptor status (Appendix 7).

Discussion

Main Findings

In this instance, our combined connectivity mapping and pharmacoepidemiological approach did not identify any additional medications which were substantially associated with breast cancer risk. Of the six medications expected to increase breast cancer risk, only verapamil had a noticeably higher risk, and this was relatively moderate in size (16%). None of four medications predicted to reduce breast cancer had meaningfully lower risk. Indeed for one of these, bendroflumethiazide, users had an 11% increased breast cancer risk compared to non-users.

Strengths and Limitations

Our study is the first to combine connectivity mapping and pharmacoepidemiology to identify medications which may alter breast cancer. The utility and robustness of our connectivity mapping process have been tested in other drug discovery and repurposing research[36, 37], while the large collection of gene expression profiles for FDA approved medications facilitated a thorough analysis of potential connections with breast cancer risk. Our pharmacoepidemiology study is based on high-quality and nationally representative CPRD data with long follow-up (>7.5 years) for most patients.[18-20] We used prescribing data collected as part of routine clinical care, in many cases, several years before the onset of breast cancer which accurately reflects GP prescribing practices and negates the risk of recall bias.

Our study had several potential weakness. It is possible that the cell line models used during the connectivity mapping process do not always translate to human body. Furthermore, our connectivity mapping considered breast cancer as a single disease, potentially diluting important signals for specific breast cancer subtypes (e.g. triple negative). Future connectivity mapping which is stratified by breast cancer subtype may lead to different results. Our pharmacoepidemiology study is observational and hence open to confounding. Although we have controlled for several of the key determinants of breast cancer risk through the matched design and analysis (e.g. age, comorbidities and GP practice), some other risk factors, including breastfeeding and ethnicity, were not available.[38, 39] Our primary analysis is based on GP records of breast cancer diagnosis. However recent evidence suggests that over 96% of GP-recorded breast cancers can be validated using other data sources[22], and our conclusions were unchanged when restricting cases to those with an NCDR-confirmed diagnosis.

We do not know if patients adhered to their prescribed medications, however our main conclusions were similar when restricting our analysis to patients who received multiple prescriptions (>365 DDDs), where non-compliance is less of a concern. Additionally, GP records do not contain data on over-the-counter medications which may have impaired our ability to accurately adjust for aspirin use (due to misclassification). Pharmacoepidemiology can only be used for commonly prescribed medications; other methodological approaches, such as cell or animal models, may be more suitable for rarer medications. Our study included a large number of medications which increased the probability of false-positive findings; therefore any 'statistically significant' results, specifically for bendroflumethiazide, should be interpreted with appropriate caution. Lastly, some of the medications explored in our analysis were rarely used within the UK general practice population. This may have led to lower power to detect differences in breast cancer risk between users and non-users, particularly in our subgroup analyses.

Comparisons with other research

To our knowledge, this is the first study to combine connectivity mapping and pharmacoepidemiology to identify medications which alter breast cancer risk. Of the medications we considered, only rosiglitazone has been previously investigated in studies of breast cancer risk, with a recent meta-analysis reporting no evidence of an association, which was in keeping with our results.[40] None of the other medications included in our study have been investigated specifically, however studies of their drug class were common, and are potentially informative. In general, these were consistent with our study findings; for example, meta-analyses of nonsteroidal anti-inflammatory drugs (meloxicam)[41], antibiotics (azithromycin)[42], selective serotonin reuptake inhibitors (citalopram, sertraline)[43] and statins (fluvastatin)[44] have all reported weak associations. Our finding of increased breast cancer risk among verapamil users is consistent with a recent meta-analysis of long-term calcium channel blocker users [45], while our estimate of 11% increased breast cancer risk among bendroflumethiazide users is similar to the results of two US-based studies of thiazide diuretics[46, 47], although a third found no association.[48] We found no studies investigating the association between rizatriptan or budesonide use (or their drug classes) and breast cancer risk.

Implications

Breast cancer places an important burden on population health, accounting for 520,000 deaths in Europe annually and €15 billion in costs.[2, 7] An improved understanding of the aetiology of breast cancer, or finding new cost-effective treatment options, could lead to better patient outcomes and financial savings. Although our combined connectivity mapping and pharmacoepidemiology

approach theoretically offers an attractive way to achieve these goals, we did not to identify any new medications which were substantially associated with breast cancer risk.

There are several potential reasons why the findings from the connectivity mapping process did not translate into observable differences in breast cancer risk among medication users. Firstly, this could reflect the weaknesses of connectivity mapping, perhaps due to a reliance on cell cultures, and difficulty interpreting connectivity scores.[13] It may be that the doses prescribed in routine clinical practice are insufficiently large for the cancer causing or protective effects identified during the connectivity mapping to become apparent. Further laboratory-based in-vitro models could be used to pre-screen an initial list of medications derived from connectivity mapping, and exclude those where the required dose is unlikely to be achieved. Our study combined gene expression information from 68 distinct datasets. Although we aimed to remove potential 'batch effects', it could be that inter-sample heterogeneity has obfuscated true differences between normal and breast cancer cells. Secondly, it is possible that the true effects of medication use are masked by the limitations of an observational study design. For example, non-differential misclassification of drug exposure (due to medication non-adherence) would be expected to attenuate our estimates towards the null effect, while residual confounding could bias our estimates in either direction.[49] However, given the consistently small associations found across the diverse range of medications included in our study, it seems unlikely that these factors are solely driving our results. Additional work exploring the causes of our null results are required to refine this methodology for future studies.

Our finding of moderately increased breast cancer risk among verapamil and bendroflumethiazide users could require further investigation. These associations are broadly consistent with the findings from previous studies of their drug classes[45-47], and plausible biological mechanisms exist for their use to increase breast cancer risk. Specifically, calcium channel blockers are known to interfere with apoptosis which could facilitate the division of cells with a malignant potential[50, 51], and thiazide diuretics may increase insulin resistance, an accepted risk factor for breast cancer.[52, 53]

Conclusions

In this instance, our combined connectivity mapping and pharmacoepidemiological approach did not identify any additional medications which were substantially associated with breast cancer risk. The cause of this is unknown, however it could be due to limitations in the connectivity mapping, such as implausible dosage requirements, or the pharmacoepidemiology, such as residual confounding.

Additional work exploring the causes of our null results are required to refine this methodology for future studies. There was some evidence that verapamil and bendroflumethiazide users had slightly higher breast cancer risk, however further confirmatory studies are required.

Disclosures

Acknowledgements: This study is based in part on data from the Clinical Practice Research Datalink obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. However, the interpretation and conclusions contained in this study are those of the author/s alone. This work was funded by a project grant from Cancer Research-UK [C37316/A18225]

Conflict of interest statement: None to declare.

Authors' contributions: CC, LM, KM, SDZ and FL conceived the study. JB conducted the analysis and drafted the initial manuscript. All authors critically revised the article for intellectual content and approved the final manuscript.

References

1. Office for National Statistics. *Cancer Registration Statistics 2014, England*. 2016 [cited 2016 11/10/2016]; Available from: http://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2014.
2. GLOBOCAN. *Estimated Cancer Incidence, Mortaility and Prevelence Worldwide in 2012*. 2012 03/11/2016].
3. Office for National Statistics. *Cancer survival in England: Patients diagnosed between 2010 and 2014 and followed up to 2015*. 2016 [cited 2016 11/10/16]; Available from: http://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalinenglandadultsdiagnosed/2010and2014andfollowedupto2015.
4. Blair, C.K., et al., *Physical inactivity and risk of poor quality of life among elderly cancer survivors compared to women without cancer: the Iowa Women's Health Study.* Journal of Cancer Survivorship, 2016. **10**(1): p. 103-112.
5. Klein, D., et al., *Long-term quality of life after breast cancer: a French registry-based controlled study.* Breast Cancer Research and Treatment, 2011. **129**(1): p. 125-134.
6. Keetharuth, A., et al., *Effects Of Cancer Treatment On Quality Of Life (ECTQOL): Final Results*. 2014, Decision Support Unit, ScHARR: Sheffield.
7. Luengo-Fernandez, R., et al., *Economic burden of cancer across the European Union: a population-based cost analysis.* The Lancet Oncology. **14**(12): p. 1165-1174.
8. Holmes, M.D. and W.Y. Chen, *Hiding in plain view: the potential for commonly used drugs to reduce breast cancer mortality.* Breast Cancer Research, 2012. **14**(2): p. 1-9.
9. Ashburn, T.T. and K.B. Thor, *Drug repositioning: identifying and developing new uses for existing drugs.* Nat Rev Drug Discov, 2004. **3**(8): p. 673-83.
10. Burn, J., et al., *Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial.* The Lancet. **378**(9809): p. 2081-2087.
11. Beral, V., *Breast cancer and hormone-replacement therapy in the Million Women Study.* Lancet, 2003. **362**(9382): p. 419-27.
12. Martin, R.M., et al., *What was the immediate impact on population health of the recent fall in hormone replacement therapy prescribing in England? Ecological study.* J Public Health (Oxf), 2010. **32**(4): p. 555-64.
13. Lamb, J., *Innovation - The Connectivity Map: a new tool for biomedical research.* Nature Reviews Cancer, 2007. **7**(1): p. 54-60.
14. Lamb, J., et al., *The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease.* Science, 2006. **313**(5795): p. 1929-1935.
15. Sirota, M., et al., *Discovery and preclinical validation of drug indications using compendia of public gene expression data.* Sci Transl Med, 2011. **3**(96): p. 96ra77.
16. Avorn , J., *In Defense of Pharmacoepidemiology — Embracing the Yin and Yang of Drug Research.* New England Journal of Medicine, 2007. **357**(22): p. 2219-2221.
17. Thillaiyampalam, G., et al., *An integrated meta-analysis approach to identifying medications with potential to alter breast cancer risk through connectivity mapping* BMC Bioinformatics, Under Review.
18. Herrett, E., et al., *Data Resource Profile: Clinical Practice Research Datalink (CPRD).* Int J Epidemiol, 2015. **44**(3): p. 827-36.
19. Herrett, E., et al., *Validation and validity of diagnoses in the General Practice Research Database: a systematic review.* Br J Clin Pharmacol, 2010. **69**(1): p. 4-14.
20. Jick, H., S.S. Jick, and L.E. Derby, *Validation of information recorded on general practitioner based computerised data resource in the United Kingdom.* Bmj, 1991. **302**(6779): p. 766-8.

21. Network, N.C.I. *National Cancer Data Repository*. 2016 [cited 2016 22/09/2016]; Available from: http://www.ncin.org.uk/collecting_and_using_data/national_cancer_data_repository/.
22. Boggon, R., et al., *Cancer recording and mortality in the General Practice Research Database and linked cancer registries.* Pharmacoepidemiol Drug Saf, 2013. **22**(2): p. 168-75.
23. Suissa, S., et al., *Time-window bias in case-control studies: statins and lung cancer.* Epidemiology, 2011. **22**(2): p. 228-31.
24. Tamim, H., A.A. Monfared, and J. LeLorier, *Application of lag-time into exposure definitions to control for protopathic bias.* Pharmacoepidemiol Drug Saf, 2007. **16**(3): p. 250-8.
25. World Health Organization. *The Anatomical Therapeutic Chemical Classification System With Defined Daily Doses*. 2016 08/11/2016]; Available from: http://www.who.int/classifications/atcddd/en/.
26. Charlson, M.E., et al., *A new method of classifying prognostic comorbidity in longitudinal studies: development and validation.* J Chronic Dis, 1987. **40**(5): p. 373-83.
27. Khan, N.F., et al., *Adaptation and validation of the Charlson Index for Read/OXMIS coded databases.* BMC Fam Pract, 2010. **11**: p. 1.
28. Nelson, H.D., et al., *Postmenopausal hormone replacement therapy - Scientific review.* Jama-Journal of the American Medical Association, 2002. **288**(7): p. 872-881.
29. Gierisch, J.M., et al., *Oral Contraceptive Use and Risk of Breast, Cervical, Colorectal, and Endometrial Cancers: A Systematic Review.* Cancer Epidemiology Biomarkers & Prevention, 2013. **22**(11): p. 1931-1943.
30. Schisterman, E.F., S.R. Cole, and R.W. Platt, *Overadjustment Bias and Unnecessary Adjustment in Epidemiologic Studies.* Epidemiology (Cambridge, Mass.), 2009. **20**(4): p. 488-495.
31. White, I.R., P. Royston, and A.M. Wood, *Multiple imputation using chained equations: Issues and guidance for practice.* Stat Med, 2011. **30**(4): p. 377-99.
32. Pokoradi, A.J., L. Iversen, and P.C. Hannaford, *Factors associated with age of onset and type of menopause in a cohort of UK women.* American Journal of Obstetrics and Gynecology, 2011. **205**(1): p. 13.
33. Mishra, G., R. Hardy, and D. Kuh, *Are the effects of risk factors for timing of menopause modified by age? Results from a British birth cohort study.* Menopause, 2007. **14**(4): p. 717-24.
34. Hirose, K., et al., *A Large-Scale, Hospital-Based Case-Central Study Of Risk-Factors Of Breast-Cancer According To Menopausal Status.* Japanese Journal of Cancer Research, 1995. **86**(2): p. 146-154.
35. Colditz, G.A., et al., *Risk Factors for Breast Cancer According to Estrogen and Progesterone Receptor Status.* Journal of the National Cancer Institute, 2004. **96**(3): p. 218-228.
36. Malcomson, B., et al., *Connectivity mapping (ssCMap) to predict A20-inducing drugs and their antiinflammatory action in cystic fibrosis.* Proc Natl Acad Sci U S A, 2016. **113**(26): p. E3725-34.
37. Wen, Q., et al., *KRAS mutant colorectal cancer gene signatures identified angiotensin II receptor blockers as potential therapies.* Oncotarget, 2017. **8**(2): p. 3206-3225.
38. Islami, F., et al., *Breastfeeding and breast cancer risk by receptor status—a systematic review and meta-analysis.* Annals of Oncology, 2015. **26**(12): p. 2398-2407.
39. Vainshtein, J., *Disparities in breast cancer incidence across racial/ethnic strata and socioeconomic status: a systematic review.* J Natl Med Assoc, 2008. **100**(7): p. 833-9.
40. Monami, M., I. Dicembrini, and E. Mannucci, *Thiazolidinediones and cancer: results of a meta-analysis of randomized clinical trials.* Acta Diabetol, 2014. **51**(1): p. 91-101.
41. Khuder, S.A. and A.B. Mutgi, *Breast cancer and NSAID use: a meta-analysis.* Br J Cancer, 2001. **84**(9): p. 1188-92.

42. Sergentanis, T.N., F. Zagouri, and G.C. Zografos, *Is antibiotic use a risk factor for breast cancer? A meta-analysis.* Pharmacoepidemiol Drug Saf, 2010. **19**(11): p. 1101-7.
43. Eom, C.S., S.M. Park, and K.H. Cho, *Use of antidepressants and the risk of breast cancer: a meta-analysis.* Breast Cancer Res Treat, 2012. **136**(3): p. 635-45.
44. Wu, Q.J., et al., *Statin use and breast cancer survival and risk: a systematic review and meta-analysis.* Oncotarget, 2015. **6**(40): p. 42988-3004.
45. Li, W., et al., *Calcium Channel Blockers and Risk of Breast Cancer: A Meta-Analysis of 17 Observational Studies.* PLoS ONE, 2014. **9**(9): p. e105801.
46. Largent, J.A., et al., *Hypertension, antihypertensive medication use, and breast cancer risk in the California Teachers Study cohort.* Cancer Causes Control, 2010. **21**(10): p. 1615-24.
47. Li, C.I., et al., *Relation between use of antihypertensive medications and risk of breast carcinoma among women ages 65-79 years.* Cancer, 2003. **98**(7): p. 1504-13.
48. Li, C.I., et al., *Use of anti-hypertensive medications and breast cancer risk among women 55–74 years of age.* JAMA internal medicine, 2013. **173**(17): p. 1629-1637.
49. Rothman, K.J., S. Greenland, and T.L. Lash. *Modern epidemiology*. 2008; Available from: http://public.eblib.com/choice/publicfullrecord.aspx?p=2032120.
50. Carson, D.A. and J.M. Ribeiro, *Apoptosis and disease.* Lancet, 1993. **341**(8855): p. 1251-4.
51. Pahor, M., et al., *Do calcium channel blockers increase the risk of cancer?* Am J Hypertens, 1996. **9**(7): p. 695-9.
52. Andersson, O.K., T. Gudbrandsson, and K. Jamerson, *Metabolic adverse effects of thiazide diuretics: the importance of normokalaemia.* J Intern Med Suppl, 1991. **735**: p. 89-96.
53. Muti, P., et al., *Markers of insulin resistance and sex steroid hormone activity in relation to breast cancer risk: a prospective analysis of abdominal adiposity, sebum production, and hirsutism (Italy).* Cancer Causes Control, 2000. **11**(8): p. 721-30.

Tables

Table 1: Patient characteristics, NCDR and GP recorded cases

	Cases	Controls
Count	45,147	45,147
Median Exposure Years (Min, Max)	7.8 (3.0,27.1)	7.8 (3.0,27.1)
Year of Diagnosis		
1995-2002	10,191 (22.6%)	10,165 (22.5%)
2003-2006	11,474 (25.4%)	11,489 (25.4%)
2007-2010	23,482 (52.0%)	23,493 (52.0%)
Mean Age (SD)	62.8 (14.1)	62.8 (14.1)
0-39	1,751 (3.9%)	1,751 (3.9%)
40-59	17,643 (39.1%)	17,643 (39.1%)
60-79	19,307 (42.8%)	19,307 (42.8%)
80+	6,446 (14.3%)	6,446 (14.3%)
Smoking		
No	28,369 (66.5%)	28,157 (66.8%)
Ex	4,482 (10.5%)	4,254 (10.1%)
Yes	9,817 (23.0%)	9,744 (23.1%)
Missing	2,479	2,992
Alcohol		
No	8,271 (21.2%)	8,742 (22.8%)
Ex	250 (0.6%)	248 (0.6%)
Yes	30,463 (78.1%)	29,379 (76.6%)
Missing	6,163	6,778
Obesity		
Normal	19,572 (49.8%)	19,643 (50.7%)
Overweight	12,448 (31.7%)	12,187 (31.4%)
Obese	7,268 (18.5%)	6,921 (17.9%)
Missing	5,859	6,396
Deprivation Quintile		
1 (Least Deprived)	10,665 (23.6%)	10,238 (22.7%)
2	9,573 (21.2%)	9,215 (20.4%)
3	8,956 (19.8%)	9,022 (20.0%)
4	8,637 (19.1%)	8,857 (19.6%)
5 (Most Deprived)	7,316 (16.2%)	7,815 (17.3%)
Mean Charlson Score (SD)	0.61 (1.08)	0.60 (1.05)
Any Comorbidity[†]	15,024 (33.3%)	14,959 (33.1%)
Chronic pulmonary disease	7,557 (16.7%)	7,542 (16.7%)
Diabetes	2,924 (6.5%)	2,727 (6.0%)
Renal disease	2,355 (5.2%)	2,312 (5.1%)
Rheumatological disease	1,531 (3.4%)	1,831 (4.1%)
Cerebrovascular disease	1,618 (3.6%)	1,651 (3.7%)
Other Medication Use		
Hormone Replacement Therapy	12,846 (28.5%)	11,816 (26.2%)
Statin	7,820 (17.3%)	7,807 (17.3%)
Aspirin	7,617 (16.9%)	7,613 (16.9%)
Oral Contraceptive	3,396 (7.5%)	2,991 (6.6%)
Metformin	1,773 (3.9%)	1,655 (3.7%)
Digoxin	1,084 (2.4%)	911 (2.0%)

[†] For brevity only the 5 most common comorbidities are listed. The full analysis included AIDS, cerebrovascular disease, chronic pulmonary disease, congestive heart disease, dementia, diabetes, diabetes with complications, ductal carcinoma in situ, hemiplegia, mild liver disease, moderate liver disease, myocardial infarction, peptic ulcer disease, peripheral vascular disease, renal disease, rheumatological disease.

Table 2: Breast cancer risk by level of medication use

Substance	NCDR and GP recorded cases				NCDR cases
	Cases	Controls	Unadjusted OR (95% CI)	Adjusted OR (95% CI)†	Adjusted OR (95% CI)
Meloxicam (Connectivity Score [CS] = 5.6)					
Never	43,543 (96.4%)	43,613 (96.6%)	Ref	Ref	Ref
Ever	1,604 (3.6%)	1,534 (3.4%)	1.05 (0.98,1.13)	1.04 (0.96,1.14)	1.12 (0.96,1.29)
1-365 DDDs‡	1,347 (3.0%)	1,307 (2.9%)	1.04 (0.96,1.12)	1.01 (0.92,1.11)	1.09 (0.92,1.28)
>365 DDDs	257 (0.6%)	227 (0.5%)	1.14 (0.95,1.36)	1.22 (1.00,1.49)	1.27 (0.90,1.80)
Azithromycin (CS=4.7)					
Never	44,813 (99.3%)	44,782 (99.2%)	Ref	Ref	Ref
Ever	334 (0.7%)	365 (0.8%)	0.91 (0.78,1.06)	0.93 (0.78,1.11)	1.04 (0.75,1.44)
1-365 DDDs	328 (0.7%)	359 (0.8%)	0.91 (0.78,1.06)	0.93 (0.78,1.11)	1.03 (0.74,1.42)
>365 DDDs	6 (0.0%)	6 (0.0%)	N/A	N/A	N/A
Rizatriptan (CS=4.6)					
Never	44,905 (99.5%)	44,904 (99.5%)	Ref	Ref	Ref
Ever	242 (0.5%)	243 (0.5%)	1.00 (0.83,1.19)	1.11 (0.90,1.36)	1.12 (0.79,1.59)
1-365 DDDs	228 (0.5%)	230 (0.5%)	0.99 (0.82,1.19)	1.07 (0.86,1.32)	1.14 (0.80,1.63)
>365 DDDs	14 (0.0%)	13 (0.0%)	N/A	N/A	N/A
Citalopram (CS=4.3)					
Never	41,293 (91.5%)	41,348 (91.6%)	Ref	Ref	Ref
Ever	3,854 (8.5%)	3,799 (8.4%)	1.02 (0.97,1.07)	1.01 (0.96,1.07)	1.06 (0.96,1.18)
1-365 DDDs	2,648 (5.9%)	2,612 (5.8%)	1.02 (0.96,1.08)	1.00 (0.94,1.07)	1.04 (0.92,1.17)
>365 DDDs	1,206 (2.7%)	1,187 (2.6%)	1.02 (0.94,1.11)	1.03 (0.94,1.13)	1.14 (0.94,1.38)
Rosiglitazone (CS=3.8)					
Never	44,893 (99.4%)	44,889 (99.4%)	Ref	Ref	Ref
Ever	254 (0.6%)	258 (0.6%)	0.98 (0.83,1.17)	1.00 (0.81,1.22)	0.88 (0.62,1.26)
1-365 DDDs	112 (0.2%)	102 (0.2%)	1.10 (0.84,1.44)	1.03 (0.76,1.41)	1.11 (0.67,1.82)
>365 DDDs	142 (0.3%)	156 (0.3%)	0.91 (0.72,1.14)	0.97 (0.75,1.26)	0.73 (0.46,1.16)
Verapamil (CS=3.6)					
Never	44,671 (98.9%)	44,753 (99.1%)	Ref	Ref	Ref
Ever	476 (1.1%)	394 (0.9%)	1.21 (1.06,1.39)	1.16 (0.99,1.37)	1.11 (0.86,1.44)
1-365 DDDs	211 (0.5%)	179 (0.4%)	1.18 (0.97,1.44)	1.08 (0.85,1.37)	1.17 (0.81,1.70)
>365 DDDs	265 (0.6%)	215 (0.5%)	1.24 (1.03,1.49)	1.24 (1.00,1.54)	1.06 (0.74,1.52)
Bendroflumethiazide (CS=-4.6)					
Never	35,783 (79.3%)	36,347 (80.5%)	Ref	Ref	Ref
Ever	9,364 (20.7%)	8,800 (19.5%)	1.09 (1.06,1.13)	1.11 (1.06,1.15)	1.10 (1.03,1.18)
1-365 DDDs	2,925 (6.5%)	2,795 (6.2%)	1.07 (1.01,1.13)	1.07 (1.01,1.15)	1.09 (0.98,1.21)
>365 DDDs	6,439 (14.3%)	6,005 (13.3%)	1.10 (1.06,1.15)	1.12 (1.07,1.18)	1.11 (1.02,1.20)
Sertraline (CS=-4.6)					
Never	43,784 (97.0%)	43,790 (97.0%)	Ref	Ref	Ref
Ever	1,363 (3.0%)	1,357 (3.0%)	1.00 (0.93,1.09)	1.01 (0.92,1.11)	1.03 (0.88,1.22)
1-365 DDDs	906 (2.0%)	899 (2.0%)	1.01 (0.92,1.11)	1.03 (0.92,1.14)	1.01 (0.83,1.23)
>365 DDDs	457 (1.0%)	458 (1.0%)	1.00 (0.88,1.14)	0.98 (0.84,1.14)	1.08 (0.82,1.43)
Fluvastatin (CS=-5.0)					
Never	44,944 (99.6%)	44,945 (99.6%)	Ref	Ref	Ref
Ever	203 (0.4%)	202 (0.4%)	1.01 (0.82,1.23)	0.98 (0.78,1.23)	1.17 (0.79,1.73)
1-365 DDDs	134 (0.3%)	125 (0.3%)	1.07 (0.84,1.38)	0.99 (0.75,1.32)	1.19 (0.74,1.92)
>365 DDDs	69 (0.2%)	77 (0.2%)	0.90 (0.65,1.24)	0.95 (0.65,1.39)	1.12 (0.56,2.23)
Budesonide (CS=-7.5)					
Never	43,578 (96.5%)	43,535 (96.4%)	Ref	Ref	Ref
Ever	1,569 (3.5%)	1,612 (3.6%)	0.97 (0.90,1.04)	0.96 (0.88,1.04)	0.90 (0.77,1.06)
1-365 DDDs	1,041 (2.3%)	1,042 (2.3%)	1.00 (0.91,1.09)	0.97 (0.87,1.07)	0.84 (0.70,1.01)
>365 DDDs	528 (1.2%)	570 (1.3%)	0.92 (0.82,1.04)	0.93 (0.81,1.07)	1.08 (0.82,1.42)

†Adjusted for comorbidities (AIDS, cerebrovascular disease, chronic pulmonary disease, congestive heart disease, dementia, diabetes, diabetes with complications, ductal carcinoma in situ, hemiplegia, mild liver disease, moderate liver disease, myocardial infarction, peptic ulcer disease, peripheral vascular disease, renal disease, rheumatological disease), confounder medications (aspirin, digoxin, hormone replacement therapy, metformin, oral contraceptive, statin), deprivation, smoking status, alcohol consumption and obesity. Additionally conditioned on age, GP practice and year of diagnosis

‡Defined daily dose

Table 3: Sensitivity analysis of breast cancer risk by ever medication use, NCDR and GP recorded cases

	Cases	Controls	Unadjusted OR (95% CI)	Adjusted OR (95% CI) †
Meloxicam				
2-year exposure lag	1,427 (3.2%)	1,361 (3.0%)	1.05 (0.97,1.14)	1.04 (0.95,1.14)
Ever use ≥3 prescriptions	680 (1.5%)	665 (1.5%)	1.02 (0.92,1.14)	1.07 (0.94,1.21)
MI lifestyle adjusted‡	1,604 (3.6%)	1,534 (3.4%)	1.05 (0.98,1.13)	1.06 (0.98,1.14)
Azithromycin				
2-year exposure lag	289 (0.6%)	310 (0.7%)	0.93 (0.79,1.09)	0.94 (0.78,1.14)
Ever use ≥3 prescriptions	40 (0.1%)	47 (0.1%)	0.85 (0.56,1.30)	0.92 (0.57,1.49)
MI lifestyle adjusted	334 (0.7%)	365 (0.8%)	0.91 (0.78,1.06)	0.90 (0.77,1.05)
Rizatriptan				
2-year exposure lag	211 (0.5%)	208 (0.5%)	1.01 (0.84,1.23)	1.09 (0.88,1.36)
Ever use ≥3 prescriptions	108 (0.2%)	131 (0.3%)	0.82 (0.64,1.06)	0.84 (0.63,1.12)
MI lifestyle adjusted	242 (0.5%)	243 (0.5%)	1.00 (0.83,1.19)	1.00 (0.83,1.20)
Citalopram				
2-year exposure lag	3,282 (7.3%)	3,253 (7.2%)	1.01 (0.96,1.06)	1.01 (0.95,1.08)
Ever use ≥3 prescriptions	2,556 (5.7%)	2,469 (5.5%)	1.04 (0.98,1.10)	1.05 (0.98,1.12)
MI lifestyle adjusted	3,854 (8.5%)	3,799 (8.4%)	1.02 (0.97,1.07)	1.01 (0.96,1.06)
Rosiglitazone				
2-year exposure lag	218 (0.5%)	223 (0.5%)	0.98 (0.81,1.18)	1.01 (0.81,1.25)
Ever use ≥3 prescriptions	215 (0.5%)	226 (0.5%)	0.95 (0.79,1.15)	0.95 (0.76,1.18)
MI lifestyle adjusted	254 (0.6%)	258 (0.6%)	0.98 (0.83,1.17)	0.92 (0.76,1.10)
Verapamil				
2-year exposure lag	427 (0.9%)	372 (0.8%)	1.15 (1.00,1.32)	1.14 (0.97,1.35)
Ever use ≥3 prescriptions	372 (0.8%)	307 (0.7%)	1.22 (1.04,1.42)	1.19 (0.99,1.43)
MI lifestyle adjusted	476 (1.1%)	394 (0.9%)	1.21 (1.06,1.39)	1.19 (1.04,1.36)
Bendroflumethiazide				
2-year exposure lag	8,647 (19.2%)	8,087 (17.9%)	1.10 (1.06,1.14)	1.11 (1.07,1.16)
Ever use ≥3 prescriptions	7,718 (17.1%)	7,204 (16.0%)	1.10 (1.06,1.14)	1.11 (1.06,1.16)
MI lifestyle adjusted	9,364 (20.7%)	8,800 (19.5%)	1.09 (1.06,1.13)	1.09 (1.05,1.13)
Sertraline				
2-year exposure lag	1,191 (2.6%)	1,179 (2.6%)	1.01 (0.93,1.10)	1.02 (0.92,1.12)
Ever use ≥3 prescriptions	826 (1.8%)	808 (1.8%)	1.02 (0.93,1.13)	1.01 (0.90,1.13)
MI lifestyle adjusted	1,363 (3.0%)	1,357 (3.0%)	1.00 (0.93,1.09)	1.00 (0.92,1.08)
Fluvastatin				
2-year exposure lag	185 (0.4%)	188 (0.4%)	0.98 (0.80,1.21)	0.95 (0.75,1.20)
Ever use ≥3 prescriptions	141 (0.3%)	155 (0.3%)	0.91 (0.72,1.14)	0.92 (0.70,1.21)
MI lifestyle adjusted	203 (0.4%)	202 (0.4%)	1.01 (0.82,1.23)	1.02 (0.83,1.25)
Budesonide				
2-year exposure lag	1,390 (3.1%)	1,449 (3.2%)	0.96 (0.89,1.03)	0.94 (0.86,1.03)
Ever use ≥3 prescriptions	997 (2.2%)	984 (2.2%)	1.01 (0.93,1.11)	0.99 (0.89,1.10)
MI lifestyle adjusted	1,569 (3.5%)	1,612 (3.6%)	0.97 (0.90,1.04)	0.96 (0.89,1.03)

†Adjusted for comorbidities (AIDS, cerebrovascular disease, chronic pulmonary disease, congestive heart disease, dementia, diabetes, diabetes with complications, ductal carcinoma in situ, hemiplegia, mild liver disease, moderate liver disease, myocardial infarction, peptic ulcer disease, peripheral vascular disease, renal disease, rheumatological disease), confounder medications (aspirin, digoxin, hormone replacement therapy, metformin, oral contraceptive, statin), deprivation, smoking status, alcohol consumption and obesity. Additionally conditioned on age, GP practice and year of diagnosis

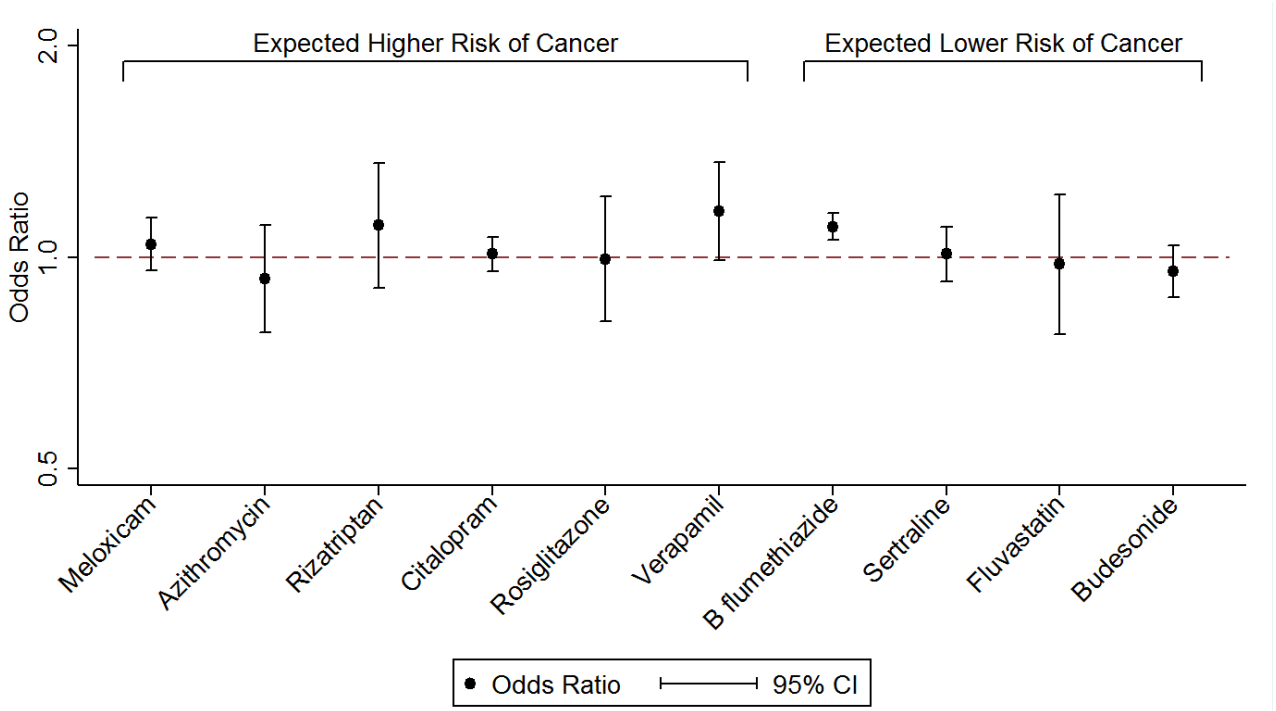
‡Multiple imputation with chained equations for smoking, alcohol consumption and obesity with age and deprivation used in the imputation, separately for cases and controls, using chained ordered logit models

Figure Legends

Figure 1: Breast cancer risk by ever medication use, NCDR and GP recorded cases. Odds ratio for each medication, ordered by decreasing connectivity score. Circle represents odds ratios, vertical lines represent the 95% confidence intervals.

Figures

Figure 1: Breast cancer risk by ever medication use, NCDR and GP recorded cases†



†Adjusted for comorbidities (AIDS, cerebrovascular disease, chronic pulmonary disease, congestive heart disease, dementia, diabetes, diabetes with complications, ductal carcinoma in situ, hemiplegia, mild liver disease, moderate liver disease, myocardial infarction, peptic ulcer disease, peripheral vascular disease, renal disease, rheumatological disease), confounder medications (aspirin, digoxin, hormone replacement therapy, metformin, oral contraceptive, statin), deprivation, smoking status, alcohol consumption and obesity. Additionally conditioned on age, GP practice and year of diagnosis

Appendices

Appendix 1: Brief description of connectivity score derivation

1. **Selection of samples:** Our search of two gene expression repositories (Gene Expression Omnibus and ArrayExpress) for 'breast cancer' resulted in 68 distinct datasets (7,530 eligible samples) across two groups:

    a. Normal: Breast tissue samples from healthy individuals (n=212)
    b. Tumour: Pre-treatment primary breast tumour samples (n=7,318)

2. **Processing of gene expression data:** The Affymetrix MAS5 algorithm was applied to generate an expression data matrix for each of the 68 datasets, and these were merged into a single file. Combining these heterogeneous datasets can lead to 'batch effects', so we used the ComBat correction method to remove these.

3. **Differential expression analysis and filtering:** We preformed differential expression analysis to identify differentially expressed genes across the tumour and normal groups. The statistical significance of any differential expression was assessed using the non-parametric two-sample Wilcoxon test. We used a p-value of 1/22277 (the number of comparisons), to account for multiple testing. The genes that passed the statistical significance filter were further examined on their magnitude of differential expression to ensure they were also biologically significant. 415 gene probes were selected as both statistically and biologically significant, and served as the input to identify medications that can potentially alter breast cancer risk.

4. **Gene signature creation:** The 415 gene probes derived from step 3 could form a single gene signature for connectivity mapping. However, for technical reasons, we split these 415 gene probes into five equal length 83-gene (sub) signatures, and ran the connectivity mapping separately for each sub-signature. We used LINCS data, which includes 83,000 reference profiles from over 1,300 FDA approved medications, as the core reference database in connectivity mapping. Each sub gene signature was used as an input to calculate the connection score (which can be interpreted similarly to a z-score from a standard normal distribution) and p-value. These indicate the strength of the relationship between the input gene signature and medication. Therefore, a positive connectivity score indicates a medication with potential cancer-causing properties, whilst a negative score signifies a medication with potential cancer-preventing properties.

5. **Connectivity mapping:** For each input gene signature, connectivity mapping analysis returns its connection score and associated p-value for all medications in the core database. We used a p-value of 1/1349 (number of FDA approved medications) to guard against multiple testing. We selected only the 67 medications which were statistically significantly associated with at least four out of the five gene signatures described above. As these medications were connected to all (or most) of the breast cancer signatures, we could have confidence that they might affect breast cancer risk.

Further details on the connectivity mapping process are available upon request.

Appendix 2: Medications excluded from the analysis

Exclusion Reason	Medication
Limited primary care use	Acepromazine, Auranofin, Azathioprine, Bacitracin, Butalbital, Cefotiam Hydrochloride, Desipramine Hydrochloride, Diloxanide Furoate, Doxylamine Succinate, Entecavir, Epinephrine, Flecainide Acetate, Isocarboxazid, Menadione, Metaraminol Bitartrate, Milnacipran, Minoxidil, Phenindione, Quinacrine, Reserpine, Sulfacetamide, Sulfafurazole, Tadalafil, Timolol, Triprolidine Hydrochloride, Tyloxapol, Valproic-Acid, Zalcitabine
Cancer treatment	Aminolevulinic Acid, Azacitidine, Chlorambucil, Cladribine, Cytarabine, Dexrazoxane, Doxorubicin, Etoposide, Gefitinib, Gemcitabine, Gemcitabine Hydrochloride, Glutamic Acid, Methotrexate, Pazopanib, Raltitrexed, Teniposide, Topotecan Hcl, Trametinib, Tretinoin, Trihydrate, Vorinostat
Topical / nasal	Beclomethasone, Bromidine, Clotrimazole, Levocabastine, Mometasone Furoate
Available over the counter	Gamma-Linolenic Acid, Ibuprofen
Known breast cancer association	Beta-Estradiol

Appendix 3: List of drug names for each exposure compound

Substance Name	Medication
Meloxicam	Meloxicam
Azithromycin	Azithromycin, Azyter, Zedbac, Zithromax
Rizatriptan	Rizatriptan, Maxalt
Citalopram	Citalopram, Cipralex, Cipramil, Escitalopram
Rosiglitazone	Rosiglitazone
Verapamil	Verapamil, Securon, Univer , Vera-Til, Verapress , Vertab , Zolvera
Bendroflumethiazide	Bendroflumethiazide, Aprinox, Neo-Naclex, Timolol
Sertraline	Sertraline, Lustral
Fluvastatin	Fluvastatin, Dorisin, Lescol, Luvinsta, Nandovar, Pinmactil
Budesonide	Budesonide, Aircort, Budeflam, Budelin, Budenofalk, DuoResp, Entocort, Pulmicort, Rhinocort, Symbicort

Appendix 4: List of generic and proprietary drug names for each confounder compound

Substance Name	Drug Name
Aspirin	Aspirin, Asasantin, Caprin, Co-codaprin, Micropirin, Migramax, Nu-Seals
Digoxin	Digoxin, Lanoxin
Hormone Replacement Therapy	Angeliq , Bedol , Climagest , Climaval , Clinorette , Conjugated oestrogens, Crinone, Cyclogest, Elleste , Estraderm , Estradiol, Estradiol , Estradot , Estragest, Estring , Estriol, Ethinylestradiol, Evorel , Femoston, Femseven, Gestone , Indivina, Kliofem , Kliovance , Lubion , Lutigest , Novofem, Nuvelle , Oestrogel , Ovestin , Premarin , Pro-Juven, Progesterone, Progynova , Prontogest , Sandrena, Serenity , Tridestra, Trisequens , Utrogestan , Vagifem , Zumenon
Metformin	Metformin, Alogliptin, Bolamyn , Competact , Diagemet , Eucreas , Glucient , Glucophage , Janumet , Jentadueto , Komboglyze , Metabet , Metformin , Sukkarto, Synjardy , Vipdomet , Vokanamet , Xigduo
Oral Contraceptive	Acondro , Aidulan , Aizea , Alenini , Alenvona , Bimizza , Brevinor , Cerazette , Cerelle, Cerelle , Cilest , Cilique , Cimizt, Cleosensa , Daylette , Desogestrel, Desogestrel , Desomono, Desorex , Dretine , Elevin , Eloine , Emerres , Erlibelle , Estradiol , Ethinylestradiol , Feanolla , Femodene , Femodette , Gedarel , Isteranda , Juliperla , Katya , Lestramyl , Levest , Levonelle, Levonorgestrel, Levonorgestrel , Lizinna , Loestrin , Lucette , Maexeni , Marvelon , Mercilon , Microgynon , Micronor , Millinette , Munalea , Nacrez , Norethisterone , Norgeston , Noriday , Norimin , Norinyl, Ovranette , Primolut , Qlaira , Rigevidon , Sofiperla , Sunya , Upostelle , Utovlan , Yacella , Yasmin , Yaz , Zelleta , Zoely
Statin	Atorvastatin, Cholib, Crestor, Dorisin, Fluvastatin, Inegy, Lescol, Lipitor, Lipostat, Luvinsta, Pinmactil, Pravastatin, Rosuvastatin, Simvador, Simvastatin, Stefluvin, Zocor

Appendix 5: Odds ratios for confounders included in 'ever use' analysis

Characteristic	Adjusted OR (95% CI)
Smoking	
No	Ref
Ex	1.04 (0.98,1.10)
Yes	1.03 (0.99,1.07)
Alcohol	
No	Ref
Ex	0.97 (0.79,1.19)
Yes	1.10 (1.06,1.15)
Obesity	
Normal	Ref
Overweight	1.02 (0.98,1.06)
Obese	1.05 (1.01,1.10)
Deprivation Quintile	
1 (Least Deprived)	Ref
2	0.91 (0.84,0.99)
3	0.82 (0.75,0.90)
4	0.69 (0.63,0.76)
5 (Most Deprived)	0.54 (0.48,0.61)
Comorbidities	
AIDS	0.97 (0.06,15.52)
Cerebrovascular disease	0.94 (0.86,1.03)
Chronic pulmonary disease	0.99 (0.95,1.03)
Congestive heart disease	1.02 (0.90,1.17)
Dementia	0.96 (0.78,1.18)
Diabetes	1.06 (0.95,1.17)
Diabetes with complications	0.88 (0.76,1.02)
DCIS	5.70 (3.46,9.42)
Hemiplegia	1.07 (0.73,1.58)
Mild liver disease	0.97 (0.74,1.27)
Mod liver disease	1.35 (0.65,2.77)
Myocardial infarction	1.05 (0.93,1.19)
Peptic ulcer disease	1.02 (0.93,1.13)
Peripheral vascular disease	0.95 (0.83,1.08)
Renal disease	1.01 (0.93,1.09)
Rheumatological disease	0.84 (0.77,0.91)
Confounder Medications	
Aspirin	1.01 (0.96,1.06)
Digoxin	1.28 (1.13,1.45)
Hormone Replacement Therapy	1.11 (1.07,1.16)
Metformin	1.04 (0.92,1.17)
Oral Contraceptive	1.14 (1.06,1.22)
Statin	0.95 (0.90,1.00)

Appendix 6: Breast cancer risk by level of medication use, NCDR recorded cases only

Substance	Cases	Controls	Unadjusted OR (95% CI)	Adjusted OR (95% CI)[†]
Meloxicam				
Never	17,565 (96.8%)	17,623 (97.1%)	Ref	Ref
Ever	588 (3.2%)	530 (2.9%)	1.12 (0.99,1.27)	1.12 (0.96,1.29)
1-365 DDDs[‡]	497 (2.7%)	447 (2.5%)	1.13 (0.98,1.29)	1.09 (0.92,1.28)
>365 DDDs	91 (0.5%)	83 (0.5%)	1.11 (0.82,1.49)	1.27 (0.90,1.80)
Azithromycin				
Never	18,045 (99.4%)	18,047 (99.4%)	Ref	Ref
Ever	108 (0.6%)	106 (0.6%)	1.02 (0.78,1.34)	1.04 (0.75,1.44)
1-365 DDDs	107 (0.6%)	106 (0.6%)	1.01 (0.77,1.33)	1.03 (0.74,1.42)
>365 DDDs	1 (0.0%)	0 (0.0%)	N/A	N/A
Rizatriptan				
Never	18,078 (99.6%)	18,077 (99.6%)	Ref	Ref
Ever	75 (0.4%)	76 (0.4%)	0.99 (0.72,1.36)	1.12 (0.79,1.59)
1-365 DDDs	74 (0.4%)	73 (0.4%)	1.01 (0.73,1.40)	1.14 (0.80,1.63)
>365 DDDs	1 (0.0%)	3 (0.0%)	N/A	N/A
Citalopram				
Never	17,019 (93.8%)	17,101 (94.2%)	Ref	Ref
Ever	1,134 (6.2%)	1,052 (5.8%)	1.09 (0.99,1.19)	1.06 (0.96,1.18)
1-365 DDDs	819 (4.5%)	758 (4.2%)	1.09 (0.98,1.21)	1.04 (0.92,1.17)
>365 DDDs	315 (1.7%)	294 (1.6%)	1.08 (0.92,1.27)	1.14 (0.94,1.38)
Rosiglitazone				
Never	18,059 (99.5%)	18,065 (99.5%)	Ref	Ref
Ever	94 (0.5%)	88 (0.5%)	1.07 (0.80,1.43)	0.88 (0.62,1.26)
1-365 DDDs	54 (0.3%)	38 (0.2%)	1.42 (0.94,2.15)	1.11 (0.67,1.82)
>365 DDDs	40 (0.2%)	50 (0.3%)	0.80 (0.53,1.21)	0.73 (0.46,1.16)
Verapamil				
Never	17,968 (99.0%)	17,991 (99.1%)	Ref	Ref
Ever	185 (1.0%)	162 (0.9%)	1.14 (0.93,1.42)	1.11 (0.86,1.44)
1-365 DDDs	94 (0.5%)	74 (0.4%)	1.27 (0.94,1.72)	1.17 (0.81,1.70)
>365 DDDs	91 (0.5%)	88 (0.5%)	1.04 (0.77,1.39)	1.06 (0.74,1.52)
Bendroflumethiazide				
Never	14,448 (79.6%)	14,665 (80.8%)	Ref	Ref
Ever	3,705 (20.4%)	3,488 (19.2%)	1.09 (1.03,1.15)	1.10 (1.03,1.18)
1-365 DDDs	1,272 (7.0%)	1,187 (6.5%)	1.10 (1.01,1.19)	1.09 (0.98,1.21)
>365 DDDs	2,433 (13.4%)	2,301 (12.7%)	1.09 (1.02,1.16)	1.11 (1.02,1.20)
Sertraline				
Never	17,714 (97.6%)	17,717 (97.6%)	Ref	Ref
Ever	439 (2.4%)	436 (2.4%)	1.01 (0.88,1.15)	1.03 (0.88,1.22)
1-365 DDDs	293 (1.6%)	296 (1.6%)	0.99 (0.84,1.17)	1.01 (0.83,1.23)
>365 DDDs	146 (0.8%)	140 (0.8%)	1.04 (0.83,1.32)	1.08 (0.82,1.43)
Fluvastatin				
Never	18,074 (99.6%)	18,088 (99.6%)	Ref	Ref
Ever	79 (0.4%)	65 (0.4%)	1.23 (0.88,1.71)	1.17 (0.79,1.73)
1-365 DDDs	54 (0.3%)	39 (0.2%)	1.40 (0.92,2.14)	1.19 (0.74,1.92)
>365 DDDs	25 (0.1%)	26 (0.1%)	0.97 (0.56,1.68)	1.12 (0.56,2.23)
Budesonide				
Never	17,657 (97.3%)	17,637 (97.2%)	Ref	Ref
Ever	496 (2.7%)	516 (2.8%)	0.96 (0.85,1.09)	0.90 (0.77,1.06)
1-365 DDDs	348 (1.9%)	367 (2.0%)	0.95 (0.81,1.10)	0.84 (0.70,1.01)
>365 DDDs	148 (0.8%)	149 (0.8%)	0.99 (0.79,1.25)	1.08 (0.82,1.42)

[†]Adjusted for comorbidities (AIDS, cerebrovascular disease, chronic pulmonary disease, congestive heart disease, dementia, diabetes, diabetes with complications, ductal carcinoma in situ, hemiplegia, mild liver disease, moderate liver disease, myocardial infarction, peptic ulcer disease, peripheral vascular disease, renal disease, rheumatological disease), confounder medications (aspirin, digoxin, hormone replacement therapy, metformin, oral contraceptive, statin), deprivation, smoking status, alcohol consumption and obesity. Additionally conditioned on age, GP practice and year of diagnosis

[‡] Defined daily dose

Appendix 7: Stratified analysis of breast cancer risk by ever medication use, NCDR and GP recorded cases

	Cases	Controls	Unadjusted OR (95% CI)	Adjusted OR (95% CI)†	Interaction Test
Meloxicam					
Menopause: age<50	92 (1.1%)	92 (1.1%)	1.00 (0.74,1.34)	0.85 (0.59,1.21)	0.186
Menopause: age>50	1,512 (4.1%)	1,442 (3.9%)	1.05 (0.98,1.14)	1.05 (0.96,1.15)	
Assumed ER Status: Negative	372 (3.5%)	347 (3.3%)	1.08 (0.93,1.26)	1.08 (0.90,1.29)	0.524
Assumed ER Status: Positive	1,232 (3.6%)	1,187 (3.4%)	1.04 (0.96,1.13)	1.03 (0.93,1.14)	
Azithromycin					
Menopause: Pre	58 (0.7%)	64 (0.8%)	0.90 (0.63,1.30)	0.96 (0.63,1.46)	0.930
Menopause: Post	276 (0.8%)	301 (0.8%)	0.91 (0.77,1.08)	0.92 (0.76,1.12)	
ER Status: Negative	75 (0.7%)	78 (0.7%)	0.96 (0.69,1.33)	0.96 (0.65,1.42)	0.907
ER Status: Positive	259 (0.7%)	287 (0.8%)	0.90 (0.75,1.07)	0.93 (0.76,1.13)	
Rizatriptan					
Menopause: Pre	73 (0.9%)	81 (1.0%)	0.90 (0.65,1.24)	1.21 (0.83,1.78)	0.603
Menopause: Post	169 (0.5%)	162 (0.4%)	1.04 (0.84,1.30)	1.08 (0.85,1.38)	
ER Status: Negative	63 (0.6%)	60 (0.6%)	1.05 (0.73,1.51)	1.11 (0.74,1.66)	0.926
ER Status: Positive	179 (0.5%)	183 (0.5%)	0.98 (0.79,1.20)	1.11 (0.88,1.41)	
Citalopram					
Menopause: Pre	846 (10.0%)	849 (10.1%)	1.00 (0.90,1.10)	0.98 (0.87,1.11)	0.431
Menopause: Post	3,008 (8.2%)	2,950 (8.0%)	1.02 (0.97,1.08)	1.02 (0.96,1.09)	
ER Status: Negative	989 (9.4%)	967 (9.2%)	1.03 (0.93,1.13)	1.02 (0.91,1.14)	0.831
ER Status: Positive	2,865 (8.3%)	2,832 (8.2%)	1.01 (0.96,1.07)	1.01 (0.94,1.08)	
Rosiglitazone					
Menopause: Pre	5 (0.1%)	12 (0.1%)	N/A	N/A	N/A‡
Menopause: Post	249 (0.7%)	246 (0.7%)	1.01 (0.85,1.21)	1.03 (0.83,1.27)	
ER Status: Negative	61 (0.6%)	60 (0.6%)	1.02 (0.71,1.45)	0.97 (0.64,1.47)	0.911
ER Status: Positive	193 (0.6%)	198 (0.6%)	0.97 (0.80,1.19)	1.00 (0.79,1.26)	
Verapamil					
Menopause: Pre	13 (0.2%)	14 (0.2%)	0.93 (0.44,1.98)	0.65 (0.26,1.61)	0.263
Menopause: Post	463 (1.3%)	380 (1.0%)	1.22 (1.07,1.40)	1.18 (1.00,1.39)	
ER Status: Negative	91 (0.9%)	83 (0.8%)	1.10 (0.81,1.48)	1.02 (0.71,1.46)	0.480
ER Status: Positive	385 (1.1%)	311 (0.9%)	1.24 (1.07,1.45)	1.20 (1.00,1.43)	
Bendroflumethiazide					
Menopause: Pre	322 (3.8%)	294 (3.5%)	1.10 (0.94,1.29)	1.23 (1.01,1.49)	0.845
Menopause: Post	9,042 (24.6%)	8,506 (23.2%)	1.09 (1.05,1.13)	1.10 (1.05,1.14)	
ER Status: Negative	1,935 (18.4%)	1,884 (17.9%)	1.04 (0.96,1.12)	1.08 (0.99,1.18)	0.541
ER Status: Positive	7,429 (21.4%)	6,916 (20.0%)	1.11 (1.07,1.15)	1.11 (1.06,1.17)	
Sertraline					
Menopause: Pre	289 (3.4%)	299 (3.5%)	0.96 (0.82,1.14)	0.98 (0.81,1.20)	0.671
Menopause: Post	1,074 (2.9%)	1,058 (2.9%)	1.02 (0.93,1.11)	1.02 (0.92,1.13)	
ER Status: Negative	318 (3.0%)	315 (3.0%)	1.01 (0.86,1.19)	1.04 (0.86,1.26)	0.813
ER Status: Positive	1,045 (3.0%)	1,042 (3.0%)	1.00 (0.92,1.10)	1.01 (0.91,1.11)	
Fluvastatin					
Menopause: Pre	0 (0.0%)	3 (0.0%)	N/A	N/A	N/A
Menopause: Post	203 (0.6%)	199 (0.5%)	1.02 (0.84,1.25)	0.99 (0.79,1.25)	
ER Status: Negative	47 (0.4%)	47 (0.4%)	1.00 (0.67,1.50)	1.19 (0.74,1.91)	0.277
ER Status: Positive	156 (0.5%)	155 (0.4%)	1.01 (0.80,1.26)	0.92 (0.71,1.19)	
Budesonide					
Menopause: Pre	216 (2.6%)	230 (2.7%)	0.94 (0.77,1.13)	0.93 (0.73,1.18)	0.389
Menopause: Post	1,353 (3.7%)	1,382 (3.8%)	0.98 (0.91,1.06)	0.96 (0.87,1.05)	
ER Status: Negative	339 (3.2%)	399 (3.8%)	0.84 (0.73,0.98)	0.83 (0.70,0.99)	0.050
ER Status: Positive	1,230 (3.5%)	1,213 (3.5%)	1.01 (0.94,1.10)	1.00 (0.90,1.10)	

†Adjusted for comorbidities (AIDS, cerebrovascular disease, chronic pulmonary disease, congestive heart disease, dementia, diabetes, diabetes with complications, ductal carcinoma in situ, hemiplegia, mild liver disease, moderate liver disease, myocardial infarction, peptic ulcer disease, peripheral vascular disease, renal disease, rheumatological disease), confounder medications (aspirin, digoxin, hormone replacement therapy, metformin, oral contraceptive, statin), deprivation, smoking status, alcohol consumption and obesity. Additionally conditioned on age, GP practice and year of diagnosis

‡Excluded due to low use of medication in subgroup