# Primary analysis of lipid regulating agents and dementia

General comment: I think you have a database and you have carried out an analysis of that database? But that’s not super clear. You should refer to what you did as either a study or an analyses etc and stick with that word every time you refer to it. Your thesis should be the whole kaboodle .

## Lay summary

Large-scale electronic health record (EHR) databases are created through the routine collection of information from patients. Under this system, whenever a patient attends their GP, their clinical data is recorded in a central database using a standard set of codes. These databases have substantial advantages for research studies, including the number of people they contain and the length of time for which participants are followed. This is particulary important when studying diseases such as dementia, which may begin to develop long before symptoms are seen.

This analysis makes use of the Clinical Practice Research Datalink (CPRD), which contains the electronic medical records of more than 3 million people from general practices across England. Using this data, I examined the effect of treatments which lower cholesterol levels, such as statins, on the risk of dementia and related outcomes.

Lipid regulating agents were found to have minimal to no effect on Alzheimer’s disease, but increased the risk of vascular dementia. This increase is very likely due to the presence of bias in the analysis despite best efforts to address it. However, this analysis still provides important sources of information which will be used in later chapters.

## Introduction

This chapter details an analysis of a large population-based electronic health record dataset to investigate the relationship between lipid regulating agent use and dementia outcomes. The analysis aims to address two important limitations of the current evidence base as identified by the systematic review presented in Chapter ??; name the two limitations

In the first instance, it explicitly examines vascular dementia as an outcome. This is particularly important to the thesis as a whole, as the systematic review identified an absence of information on the effect of lipids on dementia risk. I don’t understand this paragraph.

In the second instance, the analysis intentionally takes a different analytical approach to that most commonly used to examine the effect of statins on dementia. This approach was taken to provide a further distinct evidence source for the triangulation exercise presented in Chapter ??.

## Methods

### Study protocol

An *a priori* protocol for this study was published,1 and amendments to this are recorded in Appendix ??.2

### Data source

Previously know as the General Practice Research Database (GPRD), the Clinical Practice Research Datalink (CPRD) is a large population-based, electronic health record (EHR) database.3 The database has been collecting primary care data from participating practices across England since 1987.4,5 It contains the primary care records for more than 10 million primary care patients in England, and is broadly representative of the UK population in terms of age, sex and ethnicity.3,6

To avoid the ambiguity of interpreting free-text clinical notes and to allow for easy analysis of the resulting data, the CPRD collects data using a predefined coding system known as Read codes.7 All clinical events can be identified by a specific code. The codes use a nested approach (see Table 1), with the initial characters defining broad diagnostic topics (e.g. Eu… - Mental and behavioural disorders), while subsequent digits provide additional information on the specific condition diagnosed (e.g. Eu001 - Dementia in Alzheimer’s disease with late onset).

Table 1: Example of CPRD Read code hierarchy, showing how “Dementia in Alzheimer’s disease with late onset” is nested under the top-level of “Mental disorders”. Broad topics are specified using the initial two alpha-numeric characters of the Read code, while subsequent characters are used to define specific conditions and context.

|  |  |  |
| --- | --- | --- |
| Level | Read code | Term |
| 1 | E…. | Mental disorders |
| 2 | Eu… | Mental and behavioral disorders |
| 3 | Eu0.. | Organic mental disorder |
| 4 | Eu00. | Dementia in Alzheimer’s disease |
| 5 | Eu001 | Dementia in Alzheimer’s disease with late onset |

Using this system, lists of relevant codes for each of the index events, exposures and outcomes used in the analysis were created.

### Cohort definition

This analysis was a prospective cohort study using data from the CPRD. It included all participants registered at a participating practice between 1 January 1995 and 29 February 2016 who had a flag for “research quality” data. Records pre-dating the 1995 cutoff were included in the original CPRD extract but were excluded from the analysis as records are expected to be more complete and reliable after this date.8 Additionally, individuals with less than 12 months of continuous records prior to cohort entry were excluded, making the effective start date of the cohort the 1st January 1996.

All events of interest were identified using predetermined code lists, which are available for inspection from the archived repository accompanying this analysis (data/code availability is discussed in Section 1.5.5).

Participants were included in our study cohort if their record contained any of the following index events: a Read code for a diagnosis of hypercholesterolemia or related condition; a Read code for prescription of a lipid-regulating agent (such as statins); a total cholesterol test result of >4 mmol/L; or an LDL-c test result of >2 mmol/L.

These index events allowed us to define a population of participants who were either at risk of hypercholesterolemia, as indicated by the elevated total or LDL cholesterol test results, or had already been diagnosed with it, as indicated by a diagnostic code/related prescription. This approach was employed in an attempt to reduce confounding by indication that we would expect to observe in the full cohort, because individuals not prescribed lipid-regulating agents would likely be less healthy across a range of variables than those prescribed lipid-regulating agents, leading to a biased association been lipid-regulating agent use and dementia. Conditioning entry into the study on being either “at-risk” or already diagnosed with hypercholesterolemia attempts to mitigate this bias.

The index date for a participant was defined as the date where the first relevant code or test result was recorded on their clinical record, and participants were followed up until the earliest of: an outcome of interest; death; end of follow-up (29 February 2016); or last registration date with their GP practice. Participants were removed from our sample if they were less than 40 years of age, had less than 12 months of “research quality” data, were simultaneously prescribed more than one lipid-regulating agent (due to the difficulty of assigning these to a single exposure group), or were diagnosed with dementia before or on the date of the index event.

### Exposures

We considered seven lipid-regulating drug classes based on groupings in the British National Formulary (BNF)9, namely: statins, fibrates, bile acid sequestrants, ezetimibe, nicotinic acid groups, ezetimibe and statin (representing one treatment containing both drugs, rather than the two classes being prescribed concurrently), and omega-3 fatty acid groups.

A participant’s drug class was assigned based on their first recorded prescription, and any drug switching was ignored in an effort to mimic an intention-to-treat approach. We did however examine how often the initial drug class was stopped (defined as last prescription of the primary class being followed by at least six months of observation), added to (defined as a second drug class being prescribed before the last prescription of the initial class), or switched (defined as a second drug class being prescribed after the last prescription of the initial class).

### Outcomes

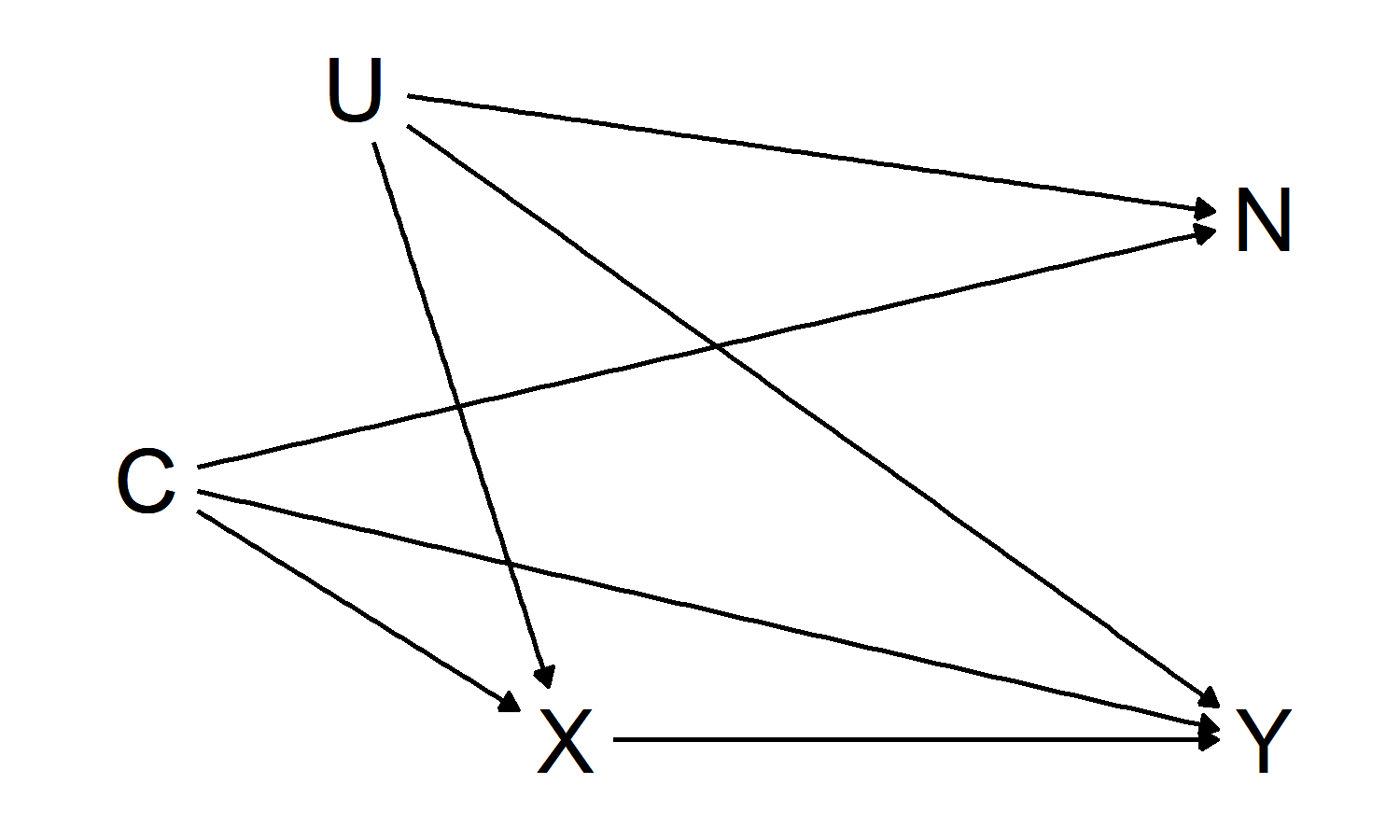
We considered five outcomes as part of this analysis: probable Alzheimer’s disease, possible Alzheimer’s disease, vascular dementia, other dementia, and a composite all-cause dementia outcome (Supplementary Figure 1). When two or more outcomes were coded in a participant’s clinical record, a decision tree was used to differentiate between them (see Figure ??). The diagnosis date of the outcome was determined by the first record of a relevant code.

Participants were followed up, and were censored at the earliest of date of diagnosis, death, date of transfer out of the study cohort or end of follow-up (29th February 2016).

In addition to our primary outcomes of interest, we created two additional control outcomes. The inclusion of control outcomes in observational analyses are a useful technique to assess the strength of uncontrolled confounding.10 Negative outcomes are those without a likely causal pathway between the exposure and outcome (see Figure @ref()). Conversely, positive control outcomes are those with a known causal association with the exposure of interest, ideally sourced from large well conducted randomised controlled trials. Positive control outcomes are useful in observational epidemiology, because if the analysis can reproduce a known result for the control outcome, our confidence in the result for the outcome of interest is increased.

Muscular backpain was used as a negative control outcome in this analysis. Despite observational analyses suggesting a link between statins and muscle pain (as opposed to more serious complications such as myopathy), systematic reviews of the adverse events of statin use11 and N-of-1 trials explicitly exploring the effect of statin use on muscle pain12 have found no evidence of an effect. As such it was decided that if a patient suffered from backpain, they should not be included in this study – is that what happened?

Additionally, incident ischemic heart disease was included as a positive control outcome, given the well-established protective effect of lipid-lowering treatment on the risk of this condition.11 This meant that patients presenting with this condition were included?!?



### Covariates

A range of additional variables were included in the analysis including X, Y and Z. These were intended to address differential distributions of potential confounding variables between those who were prescribed an LRA and those who were not.

The analysis was adjusted for a range of baseline covariates. Demographic covariates included age and gender. Age was calculated at date of entry into the cohort. Socioeconomic status was proxied using the Index of Multiple Deprivation, which draws on seven domains (income; employment; education, skills and training; health and disability; crime; barriers to housing and services; living environment) to create an overall deprivation score for each of 32844 statistical geography areas in England. Smoking and alcohol use was determined at index, and participants were categorised as current, former, or never users of each.

Body mass index (calculated as ), baseline total cholesterol and baseline LDL cholesterol measures were obtained, using the last recorded value prior to the index date. A variable indicating grouped year of entry into the cohort (<2000, 2000-2004, 2005-2009, >2010) was included to allow for changes in prescribing trends across the lifetime of the cohort. To assess healthcare utilisation, I calculated the average annual number of consultations between the beginning of a patient’s data and their entry into the cohort

Finally, presence of a range of related conditions at baseline were accounted for, including cardiovascular disease, coronary bypass surgery, coronary artery disease, peripheral arterial disease, hypertension, chronic kidney disease, and Type 1 and Type 2 diabetes. In addition to adjusting for these covariates individually, a Charlson co-morbidity index (CCI) score was calculated for each participant. The CCI is a weighted index that uses presence and severity of a number of conditions to enable adjustment for the general health of a participant in terms of their mortality risk.13 And it helped me because X.

Codelists for all covariates can be found in the archived data repository accompanying this analysis (see Section 1.5.5).

### Missing data

Given that they contain an administrative data, collected for the purposes of patient management and care rather than academic research, missing data are a recognised issue in electronic health records databases.14

In this analysis, missing data were handled using a multiple imputation approach. Variables with missing observations were identified, and 20 imputed dataset.15 Nominal variables with missing values were modelled using multinomial logistic regression, while continuous variables were modelled using linear regression. As per best practice, all variables used in the analytic model, including the outcome, were included in the imputation model.16 Imputation was performed using the MICE (Multiple Imputation by Chained Equations) command in STATA16. The result of these combined approaches was ? a clean dataset

Missing data was only considered an issue for variables where a numerical test result was expected, (e.g. BMI) or where a code existed for the absence of the condition (e.g. categorical smoking status). This approach was necessary, as absence of a code for other treatments or conditions (e.g. statin use or dementia) was assumed to indicate absence of the treatment/condition, rather than being considered missing.14

To investigate the impact of multiple imputation versus a complete case analysis,17 where participants missing any covariate are dropped from the dataset, we preformed and compared the results of both approaches.18

### Estimation methods

To estimate the effect of statins on dementia outcomes, we used a Cox proportional hazards model, defined as:

where:

* is the survival time;
* is the hazard function; and
* are the covariates which determine the hazard function, while are the coefficients for each covariate.
* is the baseline hazard - when all are zero, the function resolves to 1.

As the values of increases, value of (i.e. the hazard ratio) increases in tandem, thus reducing the survival time, .

### Time axis

As part of a Cox proportional hazard model, there is the option to use either absolute time in cohort or participants age as the time scale of interest.19–21 A model using age as the time axis inherently accounts, or adjusts, for participants age as a potential confounded of the exposure-outcome relationship. The main analyses presented all used age as the time axis.

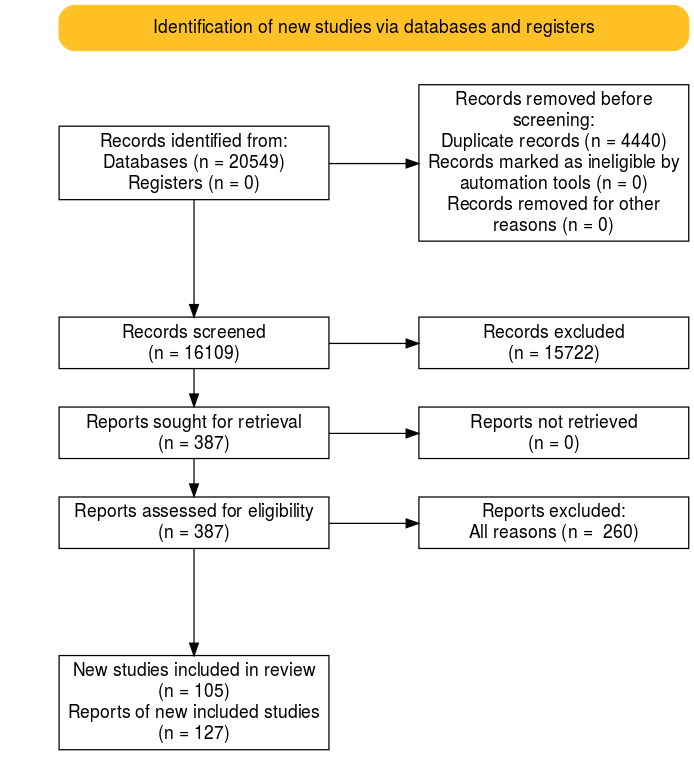
In order to explore whether the choice of time axis affects the results of the analysis,I ran two simple models, one using participants age as the time-scale and a second using absolute time-since-entry and adjusting for participant age.

### Immortal time bias and time-varying treatment indicators

Immortal time bias covers two distinct but related types of bias. The first presentation, the selection bias aspect (Panel A, Figure 2), occurs when time prior to exposure is excluded, leading to the exposed and control groups being followed up from different time points.22 In this case, events that occur prior to the exposure event are missing from the analysis.

The second presentation of immortal time bias is as a type of misclassification bias (Panel B, Figure 2). It occurs when the exposure time prior to the exposure date, and any events occurring within it, is inappropriately assigned to the exposed group. This second presentation appears to be common in the existing literature, as many of the studies included in the review in Chapter ?? were identified as being at risk of immortal time bias following formal risk of bias assessment using the ROBINS-I tool (see Section @ref(#risk-of-bias-subheading))

To address the potential for immortal time bias in our analysis, we employed a time-varying indicator of treatment status to correctly allocate time-at-risk to the exposed and unexposed groups.23



### Sensitivity analyses

To observe the effect of adjusting for additional covariates, I ran a model unadjusted except for age (captured via the time axis in the Cox model) and gender, and compared the results with the full adjusted model.

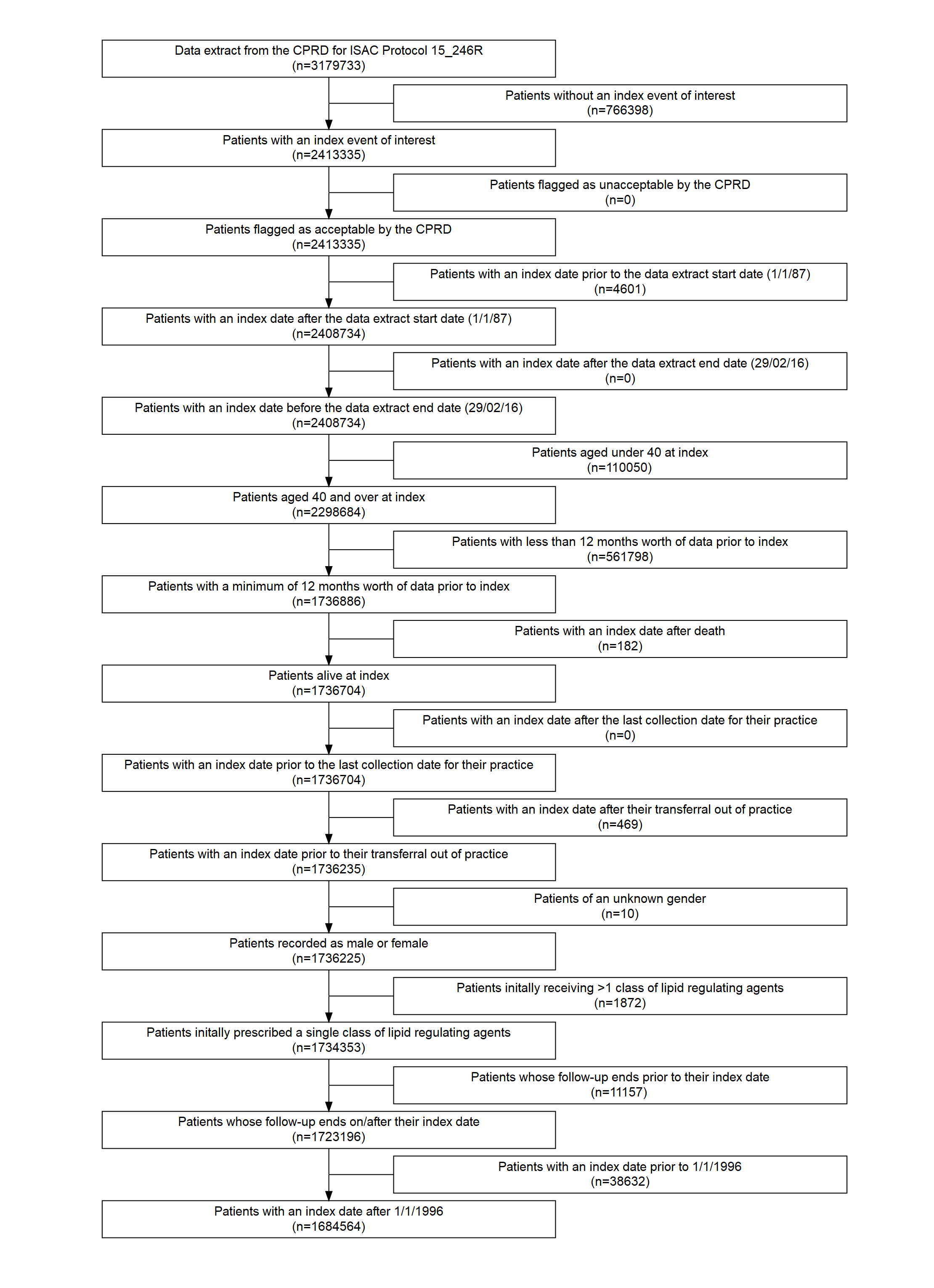
Sensitivity cohorts were also created, stratifying first by year of entry into the cohort and second by remiving participants who may have been pregnant (as statins are contraindicated in pregnancy).

Additional analyses stratified by outcome and drug class were also performed.

## Results

### Patient characteristics

Of the 3,179,733 participants included in our extract, 1,684,564 met the inclusion criteria (Figure 3), with a total follow-up of 10,800,903 patient years at risk.



The median age at index was 57 years (IQR:48-68) and participants were followed up for a median of 5.8 years (IQR:2.7-9.7). During follow-up, an all-cause dementia diagnosis was recorded for 41,830 patients (12,647 probable AD, 9,954 possible AD, 8,466 vascular dementia, 10,763 other dementia). The distribution of baseline characteristics across the drug classes can be seen in Table 2.

Table 2: Patient characteristics by drug class

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Whole Sample | None | Statins | Bile acid sequestrants | Ezetimibe | Ezetimibe & Statins | Fibrates | Nicotinic acid groups | Omega-3 Fatty Acid Groups |
| 2 | Sample size | 1695683 | 1095857 | 588455 | 5417 | 766 | 127 | 3896 | 168 | 997 |
| 3 | Index year (median) | 2006 | 2007 | 2004 | 2005 | 2004 | 2005 | 2001 | 2001 | 2005 |
| 4 | Female | 53.1% (900847) | 56.3% (616801) | 47.2% (277841) | 66.4% (3598) | 54.6% (418) | 52.8% (67) | 38.6% (1505) | 54.8% (92) | 52.7% (525) |
| 5 | Age | 57 | 54 | 62 | 57 | 60 | 57 | 58 | 62 | 56 |
| 6 | CAD | 0.4% (7180) | 0.1% (614) | 1.1% (6487) | 0.1% (6) | 0.9% (7) | 0.0% (0) | 1.4% (53) | 0.0% (0) | 1.3% (13) |
| 7 | CBS | 0.3% (5739) | 0.1% (701) | 0.8% (4947) | 0.1% (4) | 0.4% (3) | 0.0% (0) | 2.0% (78) | 0.0% (0) | 0.6% (6) |
| 8 | CVD | 2.2% (36680) | 1.2% (12803) | 4.0% (23567) | 1.7% (90) | 2.6% (20) | 2.4% (3) | 4.4% (171) | 4.8% (8) | 1.8% (18) |
| 9 | Charlson (ever > 0) | 31.0% (525566) | 25.5% (279520) | 40.9% (240925) | 42.6% (2308) | 41.8% (320) | 24.4% (31) | 50.9% (1983) | 44.6% (75) | 40.5% (404) |
| 10 | IMD-2010 (median) | 9 | 8 | 9 | 8 | 9 | 13 | 10 | 11 | 10 |
| 11 | Consulation rate (mean/SD) | 5.4 (5.5) | 5.0 (5.0) | 6.3 (6.1) | 8.6 (7.4) | 7.4 (6.6) | 4.8 (4.3) | 7.1 (6.2) | 9.3 (7.7) | 8.0 (8.0) |
| 12 | Alcohol (ever) | 85.8% (1455443) | 86.5% (947716) | 84.7% (498304) | 82.8% (4487) | 83.9% (643) | 87.4% (111) | 82.8% (3226) | 82.7% (139) | 81.9% (817) |
| 13 | Smoking (ever) | 51.1% (866207) | 47.0% (515192) | 58.5% (344540) | 55.2% (2990) | 57.4% (440) | 60.6% (77) | 60.1% (2342) | 53.6% (90) | 53.8% (536) |
| 14 | BMI (mean/SD) | 27.0 (5.3) | 26.7 (5.2) | 27.7 (5.3) | 26.8 (5.8) | 28.1 (5.7) | 28.1 (4.9) | 29.0 (5.2) | 26.4 (5.0) | 26.8 (5.5) |
| 15 | PAD | 0.8% (12879) | 0.4% (4204) | 1.4% (8524) | 0.9% (47) | 0.9% (7) | 0.8% (1) | 1.9% (75) | 6.5% (11) | 1.0% (10) |
| 16 | Hypertension | 16.1% (273011) | 11.6% (126891) | 24.5% (144010) | 12.9% (697) | 23.9% (183) | 25.2% (32) | 25.8% (1006) | 21.4% (36) | 15.6% (156) |
| 17 | Total cholesterol (mean/SD) | 5.7 (12.4) | 5.5 (10.8) | 6.2 (15.3) | 5.3 (1.3) | 7.1 (26.5) | 6.7 (1.5) | 6.4 (5.6) | 5.4 (1.5) | 5.6 (1.7) |
| 23 | LDL cholesterol (mean/SD) | 3.6 (4.8) | 3.4 (5.2) | 4.0 (3.6) | 3.1 (1.0) | 3.9 (1.1) | 4.2 (1.0) | 3.3 (1.8) | 3.4 (0.9) | 3.2 (1.0) |
| 18 | CKD | 0.1% (1391) | 0.1% (814) | 0.1% (566) | 0.1% (7) | 0.1% (1) | 0.0% (0) | 0.0% (0) | 0.0% (0) | 0.3% (3) |
| 19 | Type 1 Diabetes | 0.2% (4057) | 0.1% (795) | 0.5% (3206) | 0.3% (14) | 1.0% (8) | 0.8% (1) | 0.8% (31) | 0.6% (1) | 0.1% (1) |
| 24 | Type 2 Diabetes | 2.9% (49381) | 1.1% (12376) | 6.1% (36182) | 2.3% (124) | 5.4% (41) | 4.7% (6) | 15.8% (617) | 4.2% (7) | 2.8% (28) |

A substantial majority (98.1%) of participants prescribed a lipid-regulating agent were prescribed a statin. We excluded the “Ezetimibe and statins” and “Nicotinic acid groups” classes from subsequent analysis based on the extremely small number of participants in these groups (Table 2). The “Ezetimibe and statins” treatment group represent those prescribed a treatment containing both ezetimibe and statins, rather than those where the two treats were prescribed concurrently.

Table 3: Participants who stopped, switched or added treatements by initial treatment type

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Whole Sample | Statins | Bile acid sequestrants | Ezetimibe | Ezetimibe & Statins | Fibrates | Nicotinic acid groups | Omega-3 Fatty Acid Groups |
| 20 | Stopped | 6.8% (115899) | 19.0% (111798) | 55.9% (3028) | 19.6% (150) | 12.6% (16) | 12.3% (478) | 44.0% (74) | 35.6% (355) |
| 21 | Added | 1.6% (27470) | 4.4% (26018) | 3.5% (192) | 18.9% (145) | 3.9% (5) | 21.6% (841) | 4.2% (7) | 26.3% (262) |
| 22 | Switched | 0.9% (14956) | 2.0% (12014) | 11.3% (614) | 34.5% (264) | 64.6% (82) | 44.0% (1714) | 44.6% (75) | 19.4% (193) |

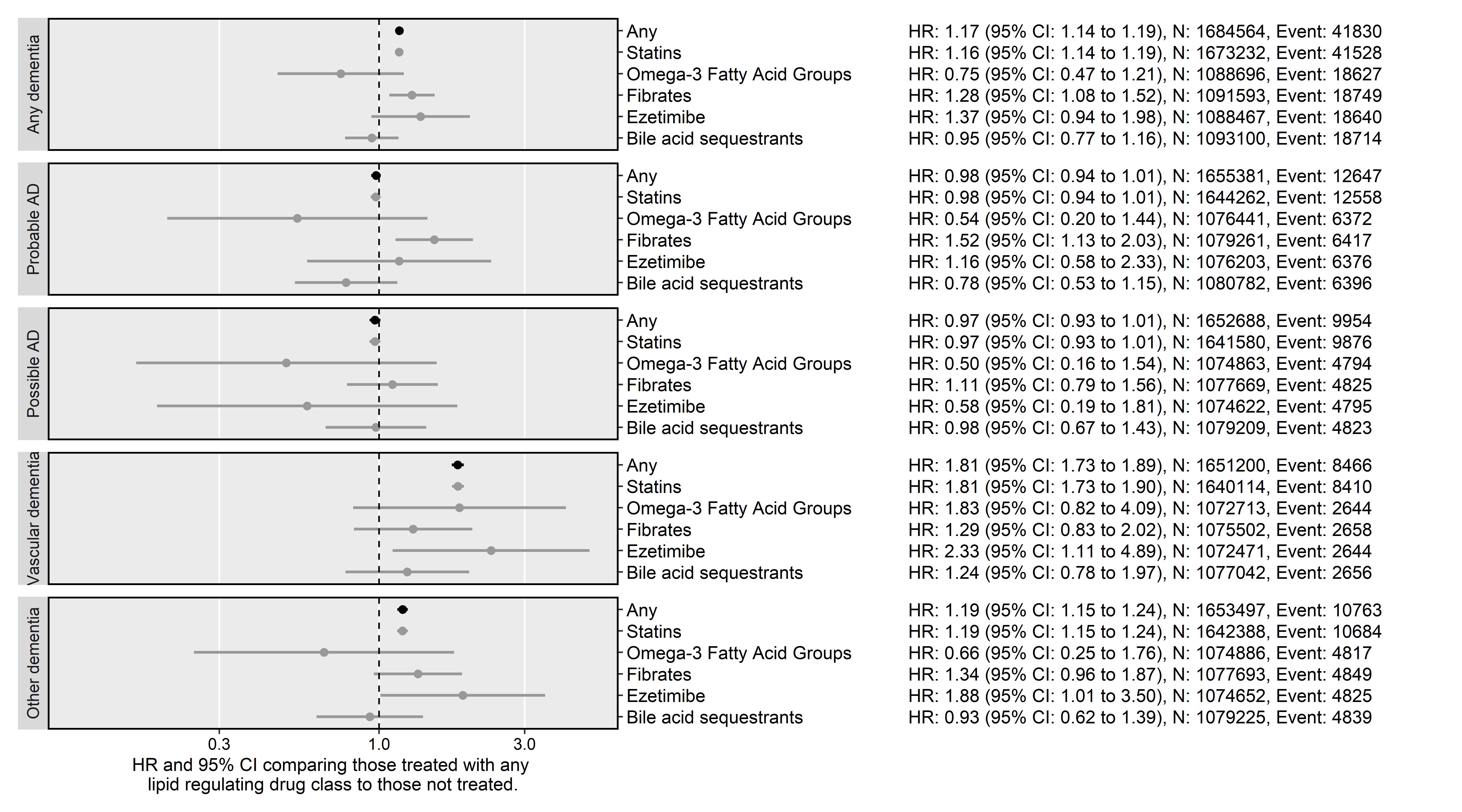
### Missing data

Full covariate information was available for 451,897 participants (26.6%). Six key variables had some missing data: IMD 2010 score, a proxy for socioeconomic position that is measured as twentiles with 1 indicating the least deprived and 20 indicating the most deprived, was missing for 630,439 participants (37.2%), because it is only recorded for English practices; alcohol status was missing for 272,745 participants (16.1%); smoking status was missing for 85,267 participants (5%); BMI, or a calculated BMI from height and weight measurements, was missing for 270,122 participants (15.9%); baseline total cholesterol was missing for 121,101 participants (7.1%), and baseline LDL cholesterol was missing for 793,720 participants (46.8%)

### Primary analysis

All results described below are from the fully adjusted model. The results of the age-only model are presented in Supplementary Figure 3, and demonstrate that adjustment for additional covariates had a minimal effect.

(ref:cprdPrimary-scap) Results from primary analyses of CPRD data



**Alzheimer’s disease**

As shown in Figure 4, the results of the analysis do not support an effect of lipid-regulating agents on probable (HR: 0.98, 95%CI: 0.94-1.01) and possible (HR: 0.97, 95%CI: 0.93-1.01) Alzheimer’s disease when compared with no treatment, with the exception of fibrates on probable Alzheimer’s disease (HR: 1.28, 95%CI: 1.08-1.52).

**Non-Alzheimer’s disease dementias**

In contrast to the findings for Alzheimer’s disease outcomes, lipid-regulating agents were associated with an increased risk of a subsequent diagnosis of vascular dementia (HR: 1.81, 95%CI: 1.73-1.89) or other dementia (HR: 1.19, 95%CI: 1.15-1.24). Again this effect was driven mainly by the statin subgroup, but there was some evidence that ezetimibe was associated with an increased risk of vascular (HR: 2.33, 95%CI: 1.11-4.89) and other (HR: 1.88, 95%CI: 1.01-3.5) dementia.

**All-cause dementia**

For the composite all-cause dementia outcome, I found treatment with a lipid-regulating agent was associated with a slightly increased risk (HR: 1.17, 95%CI: 1.14-1.19), but the magnitude of the association was not as extreme as that observed for the vascular dementia subgroup. There was also some evidence that fibrates were associated with increased risk of all-cause dementia (HR: 1.28, 95%CI: 1.08-1.52).

### Sensitivity analyses

**Choice of time axis**

**Complete case versus imputed data**

**Unadjusted versus full adjusted models**

**Sensitivity cohorts**

When stratifying based on year of entry to the cohort, I observed no variation in risk by time period in any subgroup except for probable Alzheimer’s disease (Figure ??). Removing participants coded as female and aged 55 and under at index from our analysis had minimal effect on our estimates (Figure ??).

**By statin type**

As detailed in the introduction, the properties of statins may be important in their effect, based on the ability of lip

**Control outcomes**

Using the fully adjusted model, the for the backpain negative control outcome was **PLACEHOLDER** and the ischemic heart disease was **PLACEHOLDER**

## Discussion

### Main findings

Lipid-regulating agents had a small effect on probable and possible Alzheimer’s when compared with no treatment, but were associated with increased risk of an all-cause dementia, vascular dementia and other dementia diagnosis. The effect observed in each case was driven by the statin subgroup, which included a substantial majority of participants. For the other drug classes, no association was found with any outcome, with two exceptions being that ezetimibe was associated with increased risk of vascular and other dementia, while fibrates were associated with increased risk of all-cause dementia and probable Alzheimer’s disease.

The effect estimate seems to increase as the case-mix ratio of Alzheimers:vascular dementia decreases.

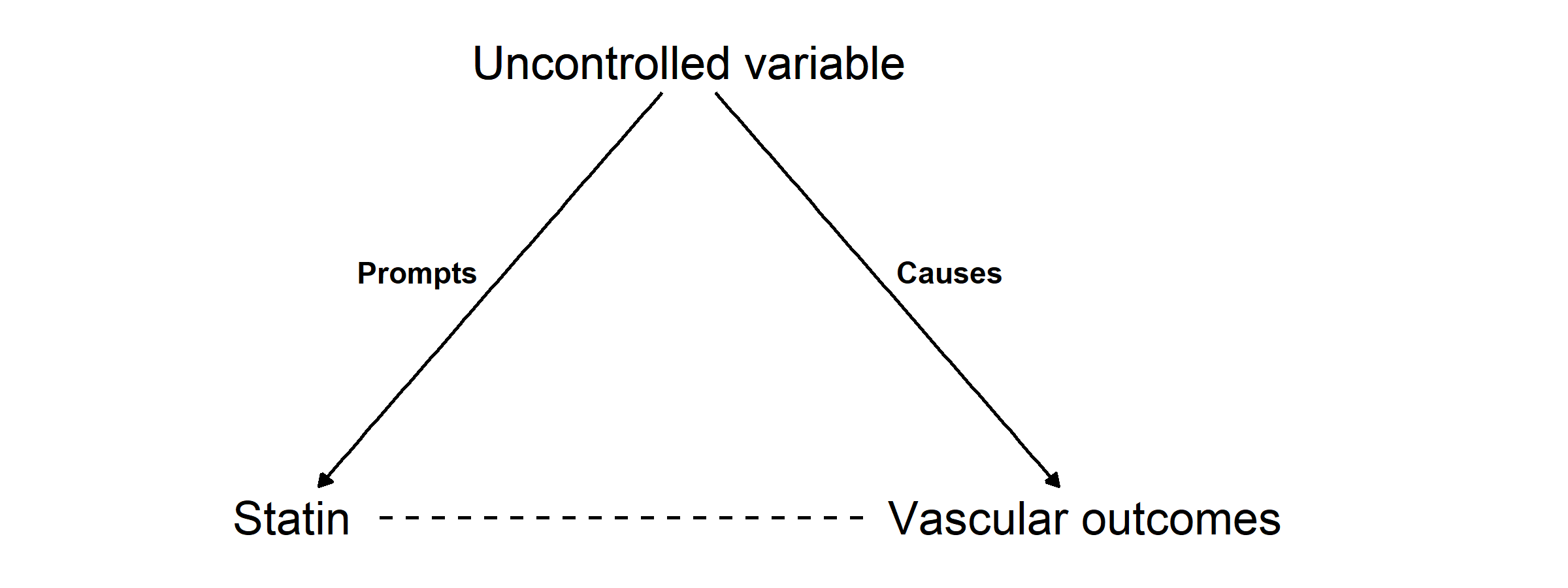
A similar increased risk of IHD is observed in those taking statins. This suggest that patients who end up taking statins (or another lipid regulating agent) are inherently different from those who do not in some way that this analysis does not account for. Is this confounding by indication? Is that why the next paragraph starts the way it does?

Confounding by indication occurs when an unobserved variable both prompts exposure to the intervention and modifies a participant’s risk of the outcome. In this case, this confounder would prompt prescription of statins (or another lipid regultating agent) but also represent a vascular risk factor that contributes to the development of the control outcome (ischemic heart disease) and also dementia outcomes with an increasing case-mix of vascular origin (e.g. the vascular dementia and other dementia endpoints).

In causal inference language, statins and dementia are said to be d-connected (see Figure 5), as there is an open “backdoor” path between them, via the uncontrolled confounders.

Rather than being a single variable, it is likely that those that take statins are different in a myriad of ways, each of which contributes a a small amount to the confounding by indication effect.

(ref:indicationBias-cap) Causal diagram (directed acyclic graph) illustrating confounding by indication



 This is still your Main Findings section – you need to wrap it up at least a little.

### Comparison to other literature

Much of the existing literature focuses on the association of statins alone with neurodegenerative outcomes, while other lipid-regulating agents are grouped as “non-statin cholesterol-lowering drugs”.24 Is this different from your study? Is this what you did too? This echoes the distribution of participants among subgroups in our analysis, with the statin subgroup including almost all participants. OK, but is there anything new?

Several previous studies have demonstrated that statins are the overwhelmingly favoured methods of cholesterol control.

You need to say something like: This section will explore comparisons in a variety of areas or something.

**Statins and all-cause dementia**

. A recent Cochrane Review carried out a study (?) on two randomized trials. These trials examined the prevention of dementia by comparing treatment with statins versus non-treatment and only one of the trials presented information on the incidence of dementia.25 This trial (Heart Protection Study) showed no effect of treatment with simvastatin on all-cause dementia risk (OR: 1.00, 95%CI:0.61-1.65),26 but concerns were raised over the diagnostic criteria used. A meta-analysis of 30 observational studies found a reduced risk of all-cause dementia was associated with statin treatment (RR 0.83, 95%CI: 0.79–0.87).27

These sources of evidence conflict with the findings of our analysis, where statin use was associated with an increased risk of all-cause dementia. However, some of the included studies in the meta-analysis specifically exclude vascular dementia from the definition of all-cause dementia,28 which may lead to an artifical protective effect of statins on all-cause dementia. So why is yours still relevant then? Two studies, diametrically opposed – why should I listen to you?

**Statins and Alzheimer’s disease**

Our results are broadly in line with the findings of two distinct approaches examining the effect of statin treatment on subsequent Alzheimer’s disease. No randomized trials of statins for the prevention of Alzheimer’s disease have been reported, but a recent meta-analysis of 20 observational studies BY WHOM? found statins were associated with a reduced risk of Alzheimer’s disease (RR 0.69, 95% CI 0.60–0.80), though the reduction was more extreme than observed in our analysis.27 In addition, a recent Mendelian randomization study examining the effect of genetic inhibition of HMGCR on Alzheimer’s disease found a small reduction in risk of Alzheimer’s disease, comparable in magnitude to our findings, but could not rule out no effect (OR: 0.91, 95%CI: 0.63-1.31).29

An additional analysis found no difference in effect between lipophilic and hydrophilic statins for the prevention of Alzheimer’s disease, consistent with a recent meta-analysis.30

Again – need some sort of “These findings are in line with the findings of this chapter” or something

**Statins and non-Alzheimer’s disease dementia**

Much less literature is available on the association between lipid-regulating agents and vascular dementia or other dementia. A recent review found four observational studies examining the association of statins and vascular dementia found evidence of an effect (RR:0.93, 95% CI 0.74–1.16).27 This contrasts with the increased effect found in our analysis. For God sake man this is a DISCUSSION. DISCUSS the fact that these are different. Draw some conclusions for me. Lead me to your point. An additional analysis found that lipophilic statins were more harmful than hydrophilic statins in vascular dementia, potentially due to their ability to cross the blood brain barrier. AND AGAIN

**Other drug classes**

Apart from statins, few studies examining a lipid-regulating agent have been reported. One of the few classes for which data was available were fibrates, which were shown to have no effect on all-cause dementia,24 inconsistent with our finding of a small increase in all-cause dementia risk in those prescribed a fibrate. I’M GOING TO HIT YOU THE NEXT TIME I SEE YOU. YOU NEED TO ACTUALLY DISCUSS THINGS!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

To our knowledge, there is no previous study of the effect of preventative treatment with ezetimibe on any dementia outcome, and so I cannot compare our unexpected finding that treatment with the drug associated with an increased risk of the vascular and other dementia outcomes.

### Strengths and limitations

There are two primary strenghts to this analysis compared to others available in the literature:

* *Size of the CPRD and length of follow-up:* Having reviewed the other studies identified by the systematic review in Chapter ??, this analysis of 1,684,564 participants is one of the largest studies of this research question.
* *Addressing the limitations of other observational analyses:* Analyzing this data has provided the opportunity to use a separate analytical technique to many of the studies identified in the systematic review. For example, the Hippsley-Cox BMJ paper examining the effect of statins, which makes use of the THIN EHR database, likely suffers from immortal time bias as exposed and unexposed participants are not followed up from a common time point.31 As touched on in the section above, this provides an additional evidence point with a different source and direction of bias, which is useful for the triangulation aspect of the thesis.

However, the findings of this analysis are subject to several limitations. There is a strong possibility of differential misclassification of dementia-related condition based on the exposure, as those with memory complaints are more likely to be classified as vascular dementia than Alzheimer’s disease if their medical records contain prescriptions for lipid-regulating agents. Further, there is a potential for non-differential misclassification of the outcome based on the use of electronic health records to identify dementia caseswhy?.32,33

This analysis may be subject to confounding by indication, which occurs when factors that affect whether a participant is exposed also affect their outcome. I attempted to address this by limiting inclusion to those either prescribed or “at risk” of being prescribed, which was determined using an elevated test result. I also adjusted for several additional potential confounding variables. However, the negative control analysis of back pain demonstrated a harmful association with lipid-regulating agent use, indicating that these findings may be biased by residual confounding. Important confounding variables for which I have not adjusted could include genetic factors maybe a second example?. A recent preprint of a study in the UK Biobank demonstrated that an Alzheimer’s disease polygenic risk score was associated with an increased risk of unspecified Alzheimer’s and vascular dementia. It was also found to be associated with an increased frequency of self-reported raised cholesterol levels, a diagnosis of hypercholesterolaemia, and a history of taking lipid-regulating agents such as statins or ezetimibe.34 This finding, combined with the potential for differential misclassification between Alzheimer’s disease and vascular dementia, could explain part of the observed association between lipid-regulating agents and vascular dementia.

Finally, there is also the potential for reverse causation in this analysis. Dementia and associated conditions have a long prodromal period, during which preclinial disease could cause indications for the prescription of a lipid-regulating agent. What have you done to counteract this or is the risk so small or unavoidableit wasn’t worth doing anything?

### Use of alternative code lists

WHen using the Smeeth et al codelists were **PLACEHOLDER** for Alzheimer’s disease and **PLACEHOLDER** for

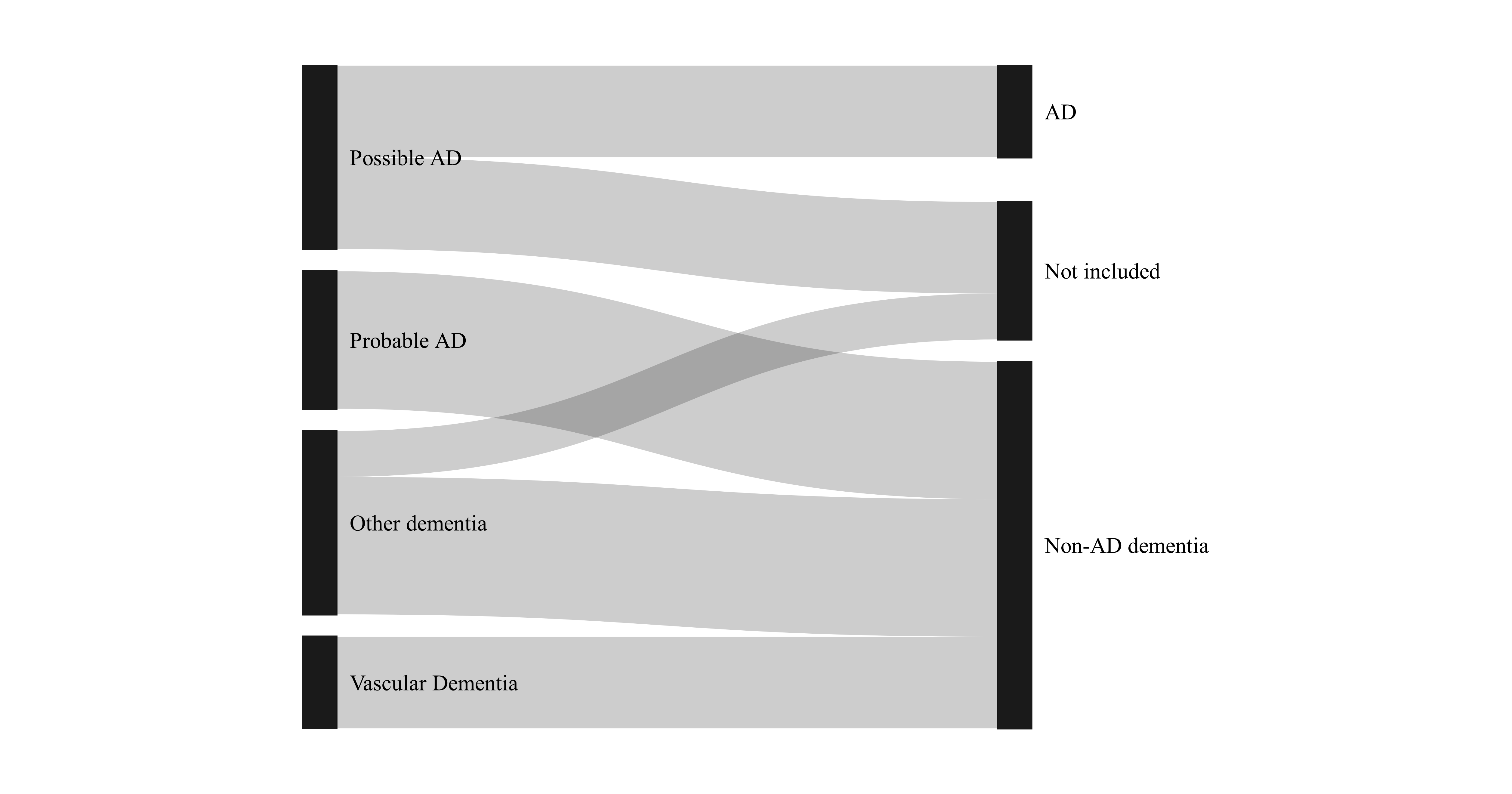
However, direct comparison between the results is difficul

Misclassification of outcomes, as discussed in the section above, is not the only issue introduced by the use of EHR codes to define outcomes. Comparing and contrasting between different studies is particularly difficult because of the impact that the use of different code lists can have on the analysis35,36 This is a particular challenge in comparing research across different time-periods and coding systems. Previous work has demonstrated that electronic health records have variable positive and negative predictive values, based substantially on the exact code-lists used to define each outcome.

As part of my exploration of the unexpected results, I attempted to replicate the analyses of another research group (Smeeth et al.)**[CITATION NEEDED]** that employed a different What is the different approach? At least vague details needed here approach to explore the effect of statin use on dementia outcomes.

However, there was a large difference in the dementia “case mix”, or proportion of different dementia diagnoses, between the two analyses, particularly in terms of the number of Alzheimer cases as a percentage of total dementia cases. The reason for this discrepancy between this analysis study and the previous paper is due to two things. While all of the codes used to define Alzheimer’s in the Smeeth paper are included in my “Probable Alzheimer’s” code-list, I included several additional codes used to define this outcome (including, for example, “Eu00013: [X]Alzheimer’s disease type 2”). Several of the codes in my “Possible Alzheimer’s” code list are included in the “Other dementia” code list used by Smeeth.

Using the Smeeth code-list, I obtained a similar breakdown of cases to that paper, with Alzheimer’s disease being a lot less frequent than when using our code-lists (which is not surprising based on the points above).



### Enabling easy synthesis of this analysis

In light of my own experiences in attempting to extract information for papers assessing preventative measures, as documented in Section ??, the outputs from this analysis are described in detail in both – in both what?.

All code, Read codelists and summary statistics (i.e. the tables presented in this chapter, plus summary tables of effect estimates) are readily available in machine readable formats (i.e. as comma separated values, or CSV, files) from the archived repository for this project.

The raw data supporting this analysis is not available, as access to the CPRD data is controlled by a data monitoring committee. Interested researchers can apply directly to the In this scenario, sharing the code represents a

As such, this analyses has been reported in line with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines37 (see Appendix ?? for the STROBE checklist).

All of these attempts at openness should aid with easy synthesis of this analysis and allow future work to be built on top (you get me)

### Conclusions

This chapter has provided new evidence on the potential repurposing of lipid-regulating agents for the prevention of all-cause dementia, Alzheimer’s disease, vascular dementia, and other dementia. It found that the use of lipid-regulating agents was associated with a small reduction in probable or possible Alzheimer’s disease, and with an increased risk of all-cause, vascular and other dementias. In all cases, the estimated associations were driven by those observed in the statin subgroup, which comprised the majority of participants in this cohort.

This chapter attempted to account for important sources of bias and provide a comparison with other available literature, as identified in the systematic review presented in Chapter ??. However, there is a strong potential for unmeasured confounding by indication and differential misclassification of the outcome on the basis of exposure, which raises questions about our findings, in particular the unexpected increase in risk of vascular dementia associated with statin use. This is supported by our findings for the negative and positive control outcomes used. Future research using large scale electronic health records should aim to address these potential biases, potentially using a analytical design that more closely emulates a trial.38

The unexpected increase in vascular dementia risk with statin use is particularly interesting given the absence of vascular dementia in the published literature, as highlighted by the previous chapter (see Section ??). It is possible that previous research identified a similar effect and via a publication bias mechanism, these results did not make it into the evidence base.

Regardless, this analysis has provided an additional source of evidence for the triangulation exercise presented in Chapter ??.

In the following Chapter, the dataset described here is incorporated along with several other datasets as part of an IPD analysis to investigate the effect of blood lipid levels on dementia outcomes directly, rather than via the proxy of treatment.

## References

1. Walker, V. M., Davies, N. M., Jones, T., Kehoe, P. G. & Martin, R. M. Can commonly prescribed drugs be repurposed for the prevention or treatment of Alzheimers and other neurodegenerative diseases? Protocol for an observational cohort study in the UK Clinical Practice Research Datalink. *BMJ Open* **6**, (2016).

2. von Elm, E. *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Journal of Clinical Epidemiology* **61**, 344–349 (2008).

3. Herrett, E. *et al.* Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology* **44**, 827–836 (2015).

4. 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. *Therapeutic Advances in Drug Safety* **3**, 89–99 (2012).

5. Wood, L. & Coulson, R. Revitalizing the general practice research database: Plans, challenges, and opportunities. *Pharmacoepidemiology and drug safety* **10**, 379–383 (2001).

6. Mathur, R. *et al.* Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *Journal of Public Health (Oxford, England)* **36**, 684–692 (2014).

7. Booth, N. What are the Read Codes? *Health Libraries Review* **11**, 177–182 (1994).

8. Wolf, A. *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *International Journal of Epidemiology* **48**, 1740–1740g (2019).

9. Wishart, D. S. *et al.* DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research* **46**, D1074–D1082 (2017).

10. Lipsitch, M., Tchetgen, E. T. & Cohen, T. Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies. *Epidemiology (Cambridge, Mass.)* **21**, 383–388 (2010).

11. Collins, R. *et al.* Interpretation of the evidence for the efficacy and safety of statin therapy. *The Lancet* **388**, 2532–2561 (2016).

12. Herrett, E. *et al.* Statin treatment and muscle symptoms: Series of randomised, placebo controlled n-of-1 trials. *BMJ* **372**, n135 (2021).

13. Charlson, M. E., Pompei, P., Ales, K. L. & MacKenzie, C. R. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of chronic diseases* **40**, 373–383 (1987).

14. Wells, B. J., Chagin, K. M., Nowacki, A. S. & Kattan, M. W. Strategies for handling missing data in electronic health record derived data. *Egems* **1**, (2013).

15. Sterne, J. A. C. *et al.* Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ* **338**, b2393 (2009).

16. Moons, K. G., Donders, R. A., Stijnen, T. & Harrell Jr, F. E. Using the outcome for imputation of missing predictor values was preferred. *Journal of clinical epidemiology* **59**, 1092–1101 (2006).

17. Pigott, T. D. A review of methods for missing data. *Educational research and evaluation* **7**, 353–383 (2001).

18. Hughes, R. A., Heron, J., Sterne, J. A. C. & Tilling, K. Accounting for missing data in statistical analyses: Multiple imputation is not always the answer. *International Journal of Epidemiology* **48**, 1294–1304 (2019).

19. Lamarca, R., Alonso, J., Gomez, G. & Munoz, A. Left-truncated Data With Age as Time Scale: An Alternative for Survival Analysis in the Elderly Population. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **53A**, M337–M343 (1998).

20. Gail, M. H., Graubard, B., Williamson, D. F. & Flegal, K. M. Comments on “Choice of time scale and its effect on significance of predictors in longitudinal studies” by Michael J. Pencina, Martin G. Larson and Ralph B. D’Agostino, *Statistics* *in* *Medicine* 2007; **26** :1343-1359. *Statistics in Medicine* **28**, 1315–1317 (2009).

21. Pencina, M. J., Larson, M. G. & D’Agostino, R. B. Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in Medicine* **26**, 1343–1359 (2007).

22. Lévesque, L. E., Hanley, J. A., Kezouh, A. & Suissa, S. Problem of immortal time bias in cohort studies: Example using statins for preventing progression of diabetes. *BMJ* **340**, b5087 (2010).

23. Lévesque, L. E., Hanley, J. A., Kezouh, A. & Suissa, S. Problem of immortal time bias in cohort studies: Example using statins for preventing progression of diabetes. *BMJ* **340**, (2010).

24. Ancelin, M.-L. *et al.* Lipid lowering agents, cognitive decline, and dementia: The three-city study. *Journal of Alzheimer’s Disease* **30**, 629–637 (2012).

25. McGuinness, B., Craig, D., Bullock, R. & Passmore, P. Statins for the prevention of dementia. *Cochrane Database of Systematic Reviews* (2016). doi:[10.1002/14651858.CD003160.pub3](https://doi.org/10.1002/14651858.CD003160.pub3)

26. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: A randomised placebocontrolled trial. *The Lancet* **360**, 7–22 (2002).

27. Poly, T. N. *et al.* Association between Use of Statin and Risk of Dementia: A Meta-Analysis of Observational Studies. *Neuroepidemiology* **54**, 214–226 (2020).

28. Chao, T.-F. *et al.* Statins and the risk of dementia in patients with atrial fibrillation: A nationwide population-based cohort study. *International Journal of Cardiology* **196**, 91–97 (2015).

29. Williams, D. M., Finan, C., Schmidt, A. F., Burgess, S. & Hingorani, A. D. Lipid lowering and Alzheimer’s disease risk: A Mendelian randomization study. *Annals of Neurology* **n/a**, (2019).

30. Chu, C.-S. *et al.* Use of statins and the risk of dementia and mild cognitive impairment: A systematic review and meta-analysis. *Scientific Reports* **8**, 5804 (2018).

31. 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* **340**, (2010).

32. Wilkinson, T. *et al.* Identifying dementia cases with routinely collected health data: A systematic review. *Alzheimer’s & Dementia* **14**, 1038–1051 (2018).

33. McGuinness, L. A., Warren-Gash, C., Moorhouse, L. R. & Thomas, S. L. The validity of dementia diagnoses in routinely collected electronic health records in the United Kingdom: A systematic review. *Pharmacoepidemiology and Drug Safety* **28**, 244–255 (2019).

34. Korologou-Linden, R. *et al.* The causes and consequences of Alzheimer’s disease: Phenome-wide evidence from Mendelian randomization. *medRxiv* 2019.12.18.19013847 (2020). doi:[10.1101/2019.12.18.19013847](https://doi.org/10.1101/2019.12.18.19013847)

35. Wilkinson, T. *et al.* Identifying dementia cases with routinely collected health data: A systematic review. *Alzheimer’s & Dementia* **14**, 1038–1051 (2018).

36. McGuinness, L. A., Warren-Gash, C., Moorhouse, L. R. & Thomas, S. L. The validity of dementia diagnoses in routinely collected electronic health records in the United Kingdom: A systematic review. *Pharmacoepidemiology and Drug Safety* **28**, 244–255 (2019).

37. Vandenbroucke, J. P. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLOS Medicine* **4**, e297 (2007).

38. Danaei, G., García Rodríguez, L. A., Cantero, O. F., Logan, R. & Hernán, M. A. Observational data for comparative effectiveness research: An emulation of randomised trials to estimate the effect of statins on primary prevention of coronary heart disease. *Statistical methods in medical research* **22**, 70–96 (2013).