When dealing with human beings controlled experiments frequently prove to be impracticable, so for a scientific basis for our assumptions we turn to past history to reconstruct the suspected causal chain of events - and then our statistical troubles may begin.

1

Primary analysis of lipid regulating agents and dementia

1.1 Additional ideas

One of the problems I'll point out with the trials in the systematic review chapter (Chapter **??**) is that they only included people with a high cardiovascular risk - we are kind of doing the same by using elevated test cholesterol results as the index event. Might be able to get around this by pointing out

1.2 Things we tried

- Conditioning entry to the cohort on QRISK score (as defined by codes) rather than. Need to be able to demonstrate here that not many meeting our original entry criteria actually go on to have a statin quickly. Need total number of those starting and average time to start.
- Positive controls. Need to be able to explain why these controls have such wildly increased results.
- Competing risk analysis, with death as a competing risk. Need numbers of deaths in each group, and preliminary results
- Allowing for time-varying confounders for binary covariates.

- Replicating a matched analysis a la Smeeth et al 2010 - though questions were raised as to how they had actually done the analysis (they did not match on propensity score, they simply adjusted for it - find paper that explores 4 ways of accounting for PS and shows this is the worst), they adjusted for things likely to be on the causal pathway, and they saw a huge change in direction for MI pre- vs post- adjustment.

1.3 Things we could try

- Work through data using positive control to see if the way it is being added is what is causing the issues.
- Examine Cramer et al. thesis and see what they did
- Marginal structural models approach (a la J. Sterne)

1.4 Aims

In this

1.5 Methods

1.5.1 Study design and protocol

We performed a prospective cohort study using data from the CPRD. Our initial sample included all participants registered at a participating practice between 1 January 1995 and 29 February 2016 who had a flag for "research quality" data. All events of interest were identified using predetermined code lists, which are available for inspection (see Data/code availability).

An *a priori* protocol for this study was published, [@walker2016a] and amendments to this are recorded in Supplementary Materials 1. This study was reported in line with the STROBE Cohort guidelines (Supplementary Table 3).[@vonelm2008]

1.5.2 Study Cohort

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 >4mmol/L; or an LDL-c test result of >2mmol/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 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 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 difficult of assigning these to a single exposure group), or were diagnosed with dementia before or on the date of the index event.

1.5.3 Exposures

We considered seven lipid-regulating drug classes based on groupings in the British National Formulary (BNF)[@wishart2017], 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).

1.5.4 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 (Supplementary Figure 1). The diagnosis date of the outcome was determined by the first record of a relevant code.

1.5.5 Covariates

The analysis was adjusted for a range of baseline covariates including sex, grouped year of entry into the cohort (<2000, 2000-2004, 2005-2009, >2010), Charlson co-morbidity index, Index of Multiple Deprivation (IMD), consultation rate, alcohol (current, former, never), smoking (current, former, never), BMI, baseline total cholesterol, and history of cardiovascular disease, coronary bypass surgery, coronary artery disease, peripheral arterial disease, hypertension, chronic kidney disease, and Type 1 and Type 2 diabetes. All covariates were determined at index and definitions for each can be found in Supplementary Table 1.

1.5.6 Estimation methods

Potential biases included time varying confounding, selection bias due to censoring on death and We use a Cox proportional hazards model with a time-varying treatment indicator.

Expand on use of age as the time scale, but describe the problems this introduced

Expand on post estimation analyses, which showed that nothing except age had an effect.

$$h(t) = h_o(t) \times exp(b_1 x_1 + b_2 x_2 + ... + b_p x_p)$$

where:

- t is the survival time;
- $h(t)$ is the hazard function; and
- $x_1, x_2, ..., x_p$ are the covariates which determine the hazard function, while $b_1, b_2, ..., b_p$ are the coefficients for each covariate.
- $h_o(t)$ is the baseline hazard - when all x_i are zero, the $exp()$ function resolves to 1.

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

Bit on why Cox models are slightly different from others, due to proportional hazards assumptions

Hazard rations are

Test for non-proportionality are therefore very important, as models which fail

Problems with testing for non-proportional hazards when

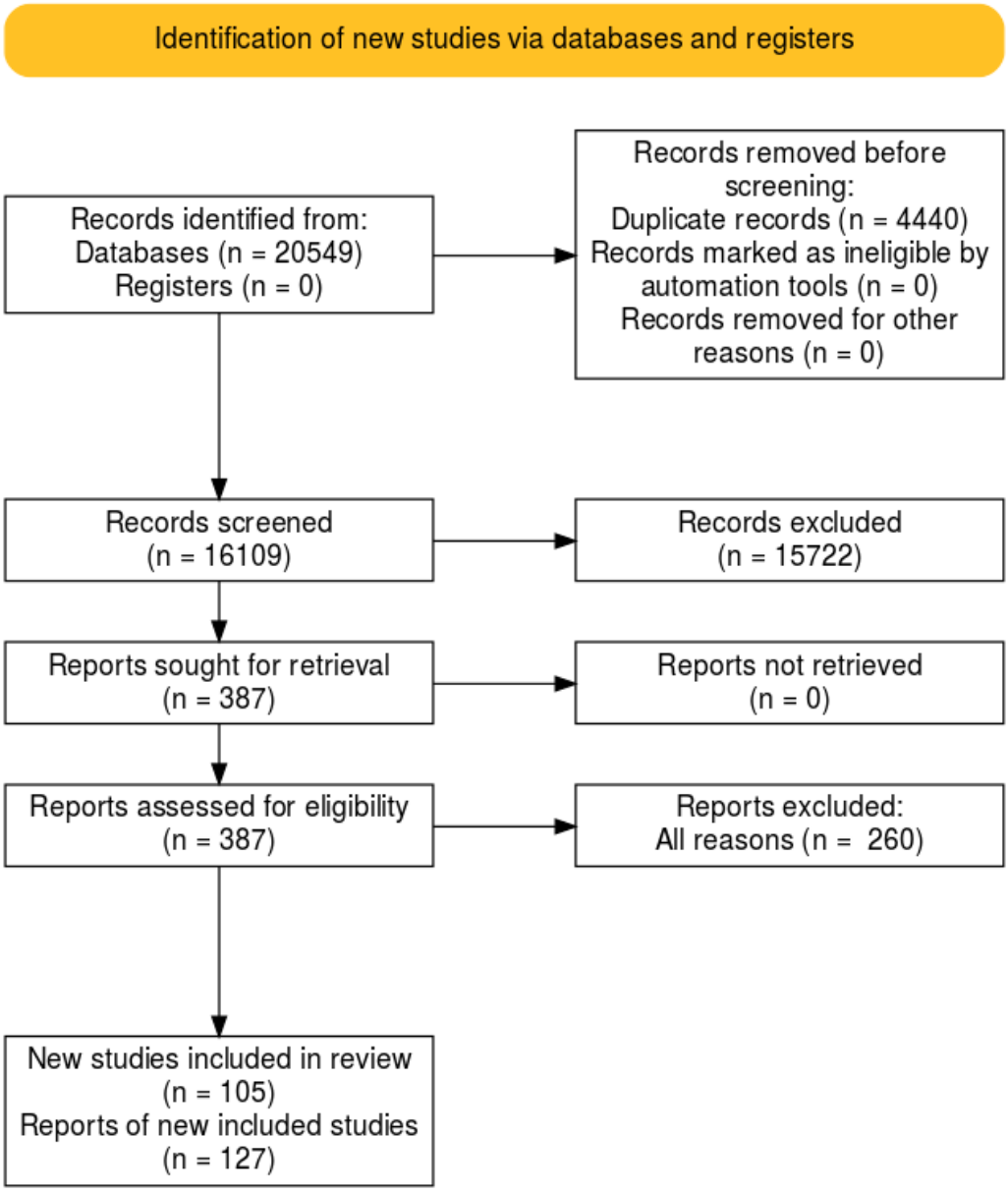

Figure 1.1: Directed acyclic graph showing assumed relationship between exposure and covariate:

1.5.7 Estimating the value of the time-varying confounders

Mean time from index event to first prescription of statins was 2.4 years. This negates the promised benefit of ruling out confounding by indication (where the test result leads to the prescription of the treatment and also increases the risk of the outcome, distorting the relationship between the two), as there is no relationship between index TC/LDL-c and eventual LRA prescription.

Additionally, the time between index event and prescription does lead to a problem in terms of time varying confounding, as an average time of 2.4 years between current measurement of the covariates and treatment switching means there is plenty of time for the value of the covariate to change. This is problematic when the decision to change treatments (in this case to move from no LRA use to LRA use) is influenced by a set of prognostic factors that in turn may have been influenced by the initial treatment decision, as is likely to be the case for a range of covariates included in the model. For example:

`_No CVD (t=0) -> No LRA (t=0) -> CVD (t=1) -> LRA (t=1) -> Dementia (t=2)_`

In this case, the decision to move to LRA use is influenced by CVD status at *Time 1*, which will not be captured by adjusting only for CVD status at *Time 0*.

In practice, this means that the value of the prognostic factor should be regularly captured

However, in electronic health records, a change in the value of the prongostic factors is only important if it is recorded in a patients record, as for it to have an impact on treatment decisions, it must be recorded.

This means we can find the most recent value of the covariate before the switch and apply a marginal structural model approach, filling all values for that variable before the most recent measure with the baseline measurement, and all after the most recent meausure with the value of the most recent measure (on the basis that you won't go from having CVD back to not having CVD).

i.e.

Timepoint 12345678

CVD 00001111

Treatment 00000111

Split into 3 month blocks since index event and use the same approach as above to work out the values of each covariate at each time point.

Note: this will be harder for things that are not dichotomous and can go up as well as down. Examples include total cholesterol and BMI, which can go up as well as down.

1.5.8 The effect of total cholesterol or LDL-cholesterol on LRA prescription

It would be fair to assume that the baseline total cholesterol/LDL-cholesterol would at least in part predict the likelihood of someone being prescribed a statin.

However, this is not the case. Baseline cholesterol level are predicted to be a poorer instrument for than QRISK2 score,[@hippisley-cox2008] which estimates a patients' 10-year risk of a cardiovascular event. Current NICE guidelines state that those with a QRISK score of 10% or higher, and in whom lifestyle modification is ineffective/inappropriate, should received a lipid regulation agent. However, this analysis could not find any effect of QRISK2 scores on statins prescription levels at 6 months.

As expected, in a confirmatory analysis using lipid levels, there was no association between *the most recent total cholesterol or LDL-cholesterol reading in the CPRD and the treatment, indicating that adjusting for this variable was not required.*

1.5.9 Replicating other analytical strategies

Comparing and contrasting between different studies is particularly difficult because of the impact that the use of different code list can have on the analysis [@wilkinson2018a; @mcguinness2019c] This highlights 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.

In order to

As part of my exploration of the unexpected results, I attempted to replicate the analyses of another paper that used a different approach to

Having contacted the authors to address these concerns, it was noted that

Propensity scores are based on a range of baseline variables, and provide an estimates of the

They used propensity scores to attempt to matcch patients equally likely to recieve the vaccine

In addition, the need to contact the authors for their code-lists reinforces the points raised in Section **??** - that all relevant material should be readily shared. Fortunately, the authors kindly shared their

1.6 Discussion

1.6.1 Main findings

Lipid-regulating agents had no 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 increase risk of all-cause dementia and probable Alzheimer's disease.

1.6.2 Comparison to other literature

Much of the existing literature focuses on the association of statins alone with neurodegenerative outcomes, with other lipid-regulating agents being grouped as "non-statin cholesterol-lowering drugs".[@ancelin2012] This echoes the distribution

of participants among subgroups in our analysis, with the statin subgroup including almost all participants.

Statins and all-cause dementia

A recent Cochrane Review identified two randomized trials comparing treatment with statins versus non-treatment for the prevention of dementia, only one of which presented information on the incidence of dementia.[@mcguinness2016a] This study (Heart Protection Study) showed no effect of treatment with simvastatin on all-cause dementia risk (OR: 1.00, 95%CI:0.61-1.65),[@heartprotectionstudycollaborativegroup2002] 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).[@poly2020]

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,[@chao2015] which may lead to a artifical protective effect of statins on all-cause dementia

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 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.[@poly2020] 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).[@williams]

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.[@chu2018]

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 no effect (RR:0.93, 95% CI 0.74–1.16).[@poly2020] This contrasts with the increased effect found in our analysis. 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.

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, [@ancelin2012] inconsistent with our finding of a small increase in all-cause dementia risk in those prescribed a fibrate.

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

1.6.3 Original contribution to research

There are two primary ways in which this research adds to the topic:

- *Size of the CPRD and length of follow-up:* Having reviewed the other studies identified by the systematic review in Chapter 4, this analysis of 1.7 million 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, As an 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.[@hippisley-cox2010] 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.

1.6.4 Strengths and Limitations

A major strength of our analysis is the size of the included cohort and the length of follow-up that the use of electronic health records allowed. In addition, we followed users and non-users from a common index date, using a time-updating treatment indicator to correctly assign time-at-risk to the exposed and unexposed groups.

However, the findings of our 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 contains 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 cases.[@wilkinson2018; @mcguinness2019b]

Our study may be subject to confounding by indication, which occurs when factors that affect whether a participant is exposed also affect their outcome. We 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. We 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, indicting that our findings may be biased by residual confounding. Important confounding variables for which we have not adjusted could include genetic factors.

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, and also 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.[@korologou-linden2020] 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.

1.6.5 Enabling easy synthesis

In light of my own experiences in attempting to extract information for papers assessing prevantive measures, I attempted to document

Study was reported in line with the [CHECK] reporting guidelines, and

1.6.6 Conclusions

We have 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. We found use of lipid-regulating agents not associated with probable or possible Alzheimer's disease, but were associated 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 our cohort.

We have attempted to account for important sources of bias in our analysis and provide a comparison with other available literature. However, there is a strong potential for unmeasured confounding, misclassification and reverse causation, which raises questions about our findings, in particular the unexpected increase in risk of vascular dementia associated with statin use. Future research should

aim to address these potential biases and, while it may be costly in terms of time and resources, a large scale, long-term randomized controlled trial would provide useful additional information on the effect of lipid-regulating agents on the risk of dementia and related outcomes.