# Annual Review - Year 2

**Title**: Investigating the relationship between blood lipids levels, and interventions that modify blood lipid levels, and dementia outcomes.

**Supervisors**:

* Julian Higgins (Primary)
* Yoav Ben-Shlomo
* Liz Coulthard
* George Davey-Smith

**PhD start date**: 01/01/2019

**Expected submission date**: 01/01/2022

**Maximum submission date**: 01/01/2023

**Note:** The content of this report represents a more comprehensive version of that entered into the STaR University of Bristol PGR reporting system.

# Overview of PhD

## Aim

The central aim of this thesis is to:

* **To investigate the relationship between blood lipid levels, and interventions that modify blood lipid levels, and dementia outcomes**

## Objectives

To achieve this aim, there are several smaller objectives, which map to the proposed chapters of my thesis outlined in Section 2:

1. Identify all existing evidence on the relationship between blood lipid levels (and interventions that affect blood lipid levels) and dementia, regardless of study design and including studies available solely as a preprint.
2. Examine the relationship between lipid-regulating agents and dementia in a large scale population-based cohort, the Clinical Practice Research Datalink.
3. Examine the relationship between blood lipid levels and dementia risk in several cohorts as part of a individual participant data meta-analysis.
4. Integrate the existing evidence identified by the review with the primary evidence produced as part of this thesis in a qualitative triangulation exercise.

## Theoretical framework

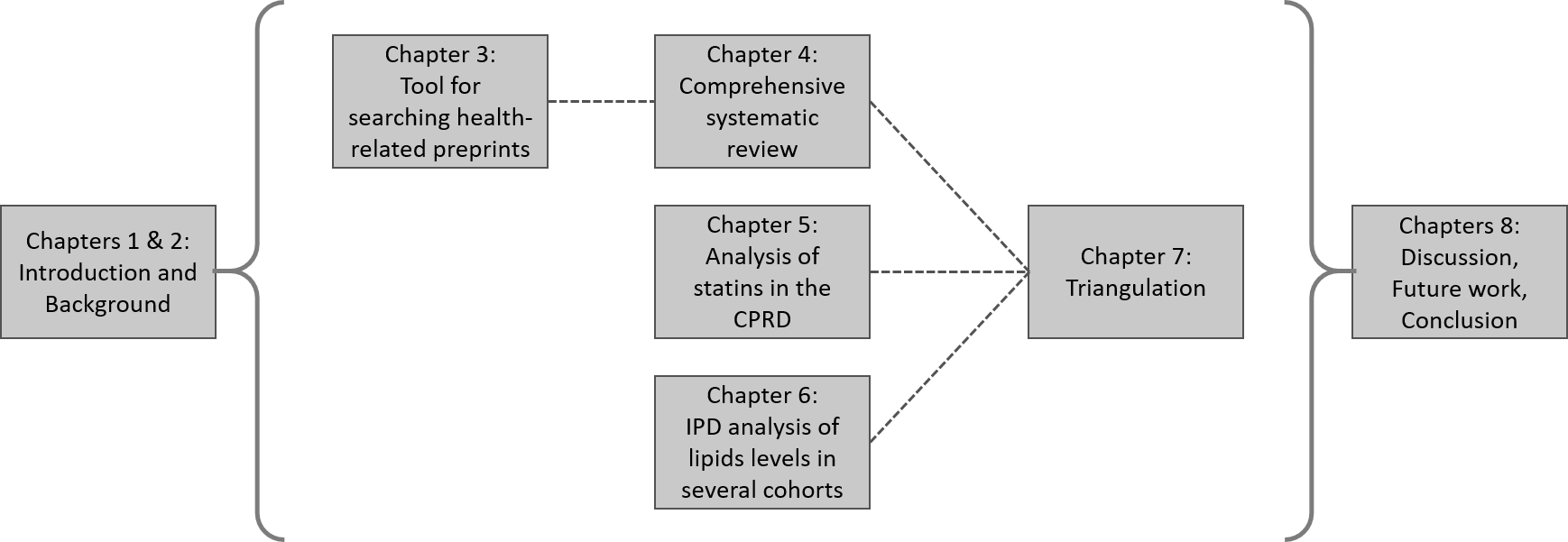
The main theoretical framework used in this thesis is evidence synthesis - the identification, critical assessment and integration of all available evidence on a research question in order to either: a) provide a more definitive answer to that question; or b) highlight gaps in the existing evidence base.

While this thesis does include a primary evidence generation element (Chapter 5 - observational analysis of CPRD data), this was performed with the intention of providing a further source of evidence for the evidence synthesis/triangulation aspects.

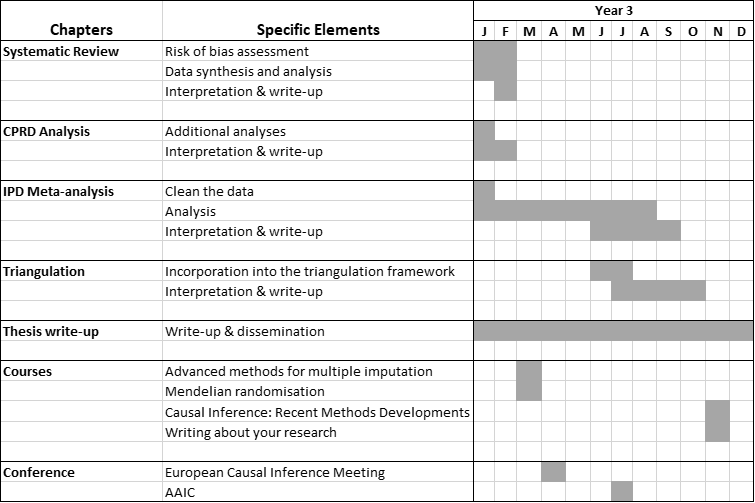
# Chapter progress

The thesis is laid out in a similar manner to that successfully defended by [Sam Abbott of the MRC IEU](https://www.samabbott.co.uk/thesis). Each chapter is self-contained, presenting the relevant methods and results for that particular analysis, and so there are no dedicated methods chapters.

An overview of the proposed thesis chapters and the way they link together is shown in Figure 1. The exact order in which chapters should be presented (particularly Chapters 4, 5 & 6) is yet to be confirmed. Any advice from reviewers on this would be appreciated.



**Figure 1: Overview of thesis chapters and the relationship between them.**



**Figure 2: Gantt chart showing research milestones, training courses and conferences for the last 12 months of my PhD thesis.**

## Chapter 1: Introduction

**Overview**

This short chapter presents the aims and objectives of the thesis, an overview of each chapter, an introduction to evidence synthesis as the theoretical framework underlying the thesis, and a summary of thesis outputs (papers/software).

**Progress**

A preliminary draft of this chapter is complete. Additional outputs will be added if/when they arise.

## Chapter 2: Background

**Overview**

This chapter provides an overview of the main topics covered in the thesis, including dementia (prevalence and impact, clinical presentation, diagnostic criteria, treatments), blood lipids fractions (fractions of interest (TC, HDL, LDL, TG), accepted ranges for each), and lipid regulating agents (type (statin vs non-statin), mechanism of action, indications for use).

**Progress**

This descriptive chapter is approximately 70% complete.

## Chapter 3: Tool for systematically searching health-related preprints (medrxivr)

**Overview**

This chapter introduces a tool built in R to allow for systematic searching of the [medRxiv](https://www.medrxiv.org/) and [bioRxiv](https://www.biorxiv.org/) preprint repositories. Preprints represent an important source of grey literature and so should be included as part of a systematic review. However, the native search interface on preprint websites do not allow for complex search strategies (e.g. those using Boolean operators or wildcards) and are not transparent or reproducible, with the same search producing wildly different numbers of records over time.

To allow for easy systematic searching of this literature source as part of the systematic review described in Chapter 4, the [medrxivr](https://docs.ropensci.org/medrxivr/) tool was developed.

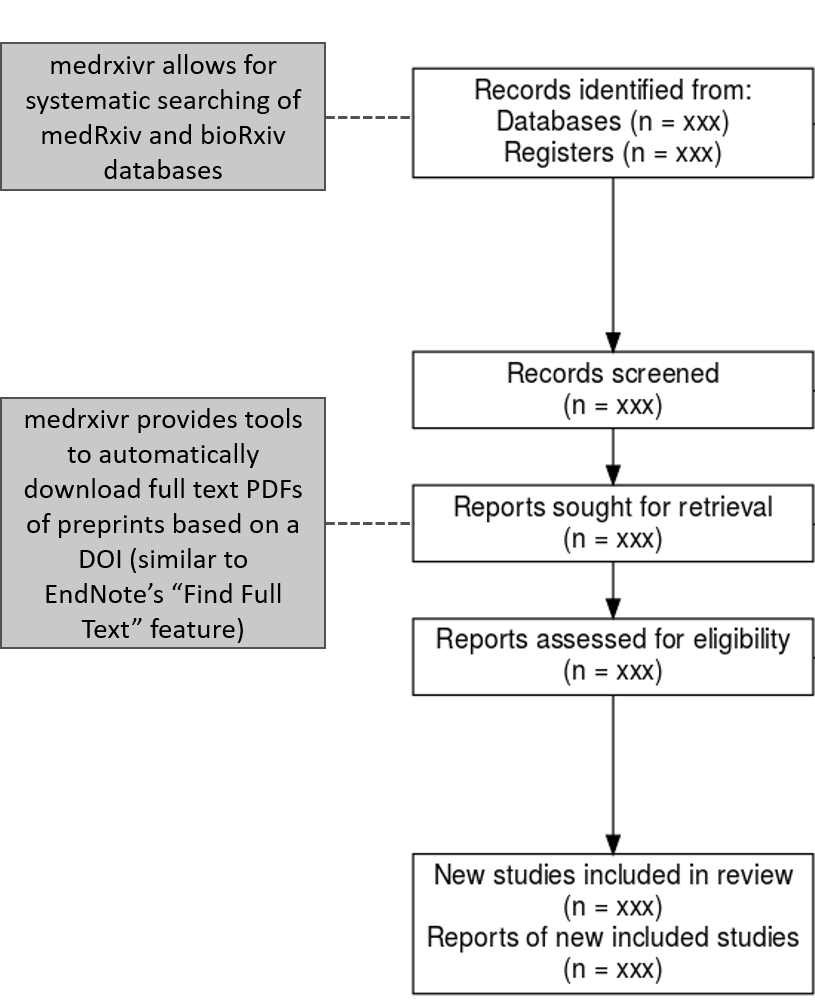
**Note:** Evidence synthesis software development has been a key component of my work during my PhD. At my last progress review, Tom Gaunt (PGR Director) flagged in his comments that I should consider including some of this work as a chapter in my thesis, if I could tie it in with the PhD as whole. Based on this, I am keen to include the medrxivr tool as it was designed and developed specifically to help with the review described in Chapter 4. However, I would be very glad to hear your opinion on whether the inclusion of this work as a chapter makes sense.

**Progress**

* Paper: 100% (published)
* Chapter: 100% (draft)

The code underlying the tool underwent peer review via the rOpenSci review process (you can see the discussion between myself and the reviewers [here](https://github.com/ropensci/software-review/issues/380)), following which a short paper describing the tool was [published in the Journal of Open Source Software](https://joss.theoj.org/papers/10.21105/joss.02651).

This chapter represents an expanded version of the published paper, discussing the motivation, key functionality, reception and planned future development of the tool. A draft of this chapter is complete.



**Figure 3 - Role of medrxivr in the systematic review process:** medrxivr contributes to the review process at two key stages: (i) searching of literature sources, and (ii) accessing full text records.

## Chapter 4: Systematic review

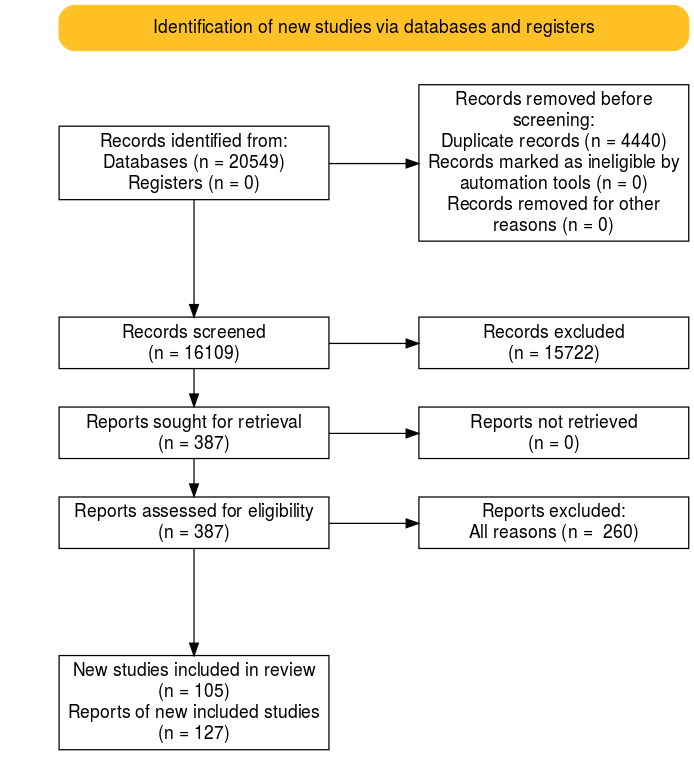
**Overview**

This chapter will present the methods and findings of a comprehensive systematic review into the relationship between blood lipids levels/blood lipids interventions (e.g. statins) and dementia. The review included all types of study design including randomised controlled trials, Mendelian randomisation analyses, and non-randomised studies of exposures and interventions (NRSI/NRSE).

**Progress**

In terms of the different stages of the review, progress currently stands as:

* Literature search: 100%
* Title and abstract/Full-text screening: 100%
* Data extraction: 100%
* Risk of bias assessment: 75%
* Analysis: 60%
* Writing: 50%

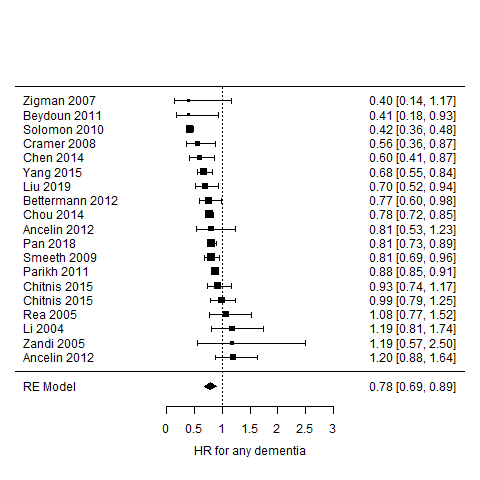


**Figure 4 - PRISMA flow diagram showing the movement of records through the systematic review.**

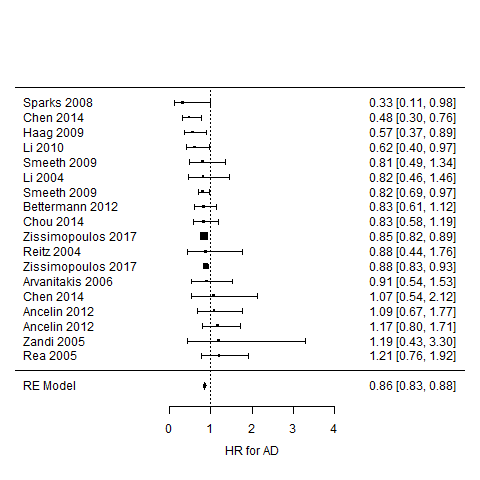
First, in response to issues flagged at last year’s progress review (comprehensiveness of search/is screening a 10% sample in duplicate sufficient if inter-rater reliability is low), searching the bibliographies of published reviews on this topic demonstrated that the search was sufficiently comprehensive to capture all previously identified studies in addition some studies that had not been included in any previous review. Additionally, inter-rater agreement on a random 10% subsample of records was high for both the title/abstract and full-text screening stages of the review.

For the sake of space, the full methods and results of the review are not presented here, but the protocol is [available online](https://doi.org/10.17605/OSF.IO/VTW5Y). As a short summary, 127 reports describing 105 primary studies were included in the review (see the PRISMA flow diagram above for an overview of the movement of records through the review). The majority of studies identified by the review are non-randomised studies, with only two randomised controlled trails and eight Mendelian randomisation studies identified. Data extraction has been completed for all studies, and risk-of-bias assessments for each are almost complete.

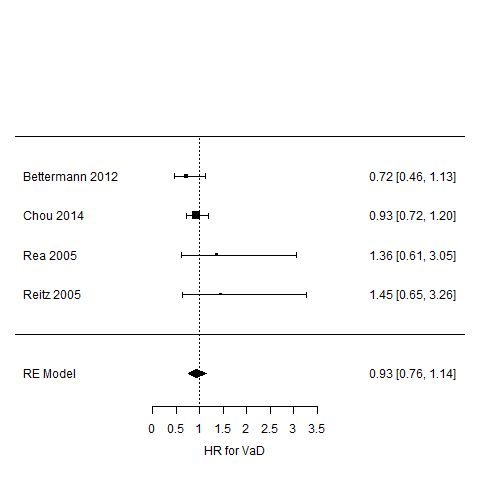
To enable comparison with the results of the observational analysis presented as part of Chapter 5, the results of a meta-analysis examining the effect of statins on dementia outcomes in included NRSI is presented. Statin use was associated with a reduced risk of all-cause dementia (HR: 0.86, 95%CI 0.83-0.88) and Alzheimer’s disease (HR: 0.78, 95%CI 0.69-0.89), but were not found to have an effect on the risk of vascular dementia (HR: 0.93, 95%CI 0.76-1.14).



**Figure 5 - Results of a random-effects meta-analysis of non-randomised studies examining the effect of statin use on the risk of all-cause dementia.**



**Figure 6 - Results of a random-effects meta-analysis of non-randomised studies examining the effect of statin use on the risk of Alzheimer’s disease.**



**Figure 7 - Results of a random-effects meta-analysis of non-randomised studies examining the effect of statin use on the risk of vascular dementia.**

In terms of the meta-analysis of the non-intervention studies, this analysis was complicated by the different ways in which lipid fractions were quantified in different studies (e.g. grouped based on clinical guidelines or quantiles, or dichotomised into a hypercholesterolemia variable). However, following discussion with my supervisors who pointed me towards useful material on this subject, this element is almost complete.

One interesting discussion point highlighted by the review is the absence of data on vascular dementia as an outcome. As illustrated in the meta-analysis above, there were substantially less non-randomised studies reporting vascular dementia as an outcome compared with those reporting all-cause dementia or Alzheimer’s disease outcomes. This is reflected in the available randomised controlled trials and Mendelian randomisation studies (as no genetic consortia data are available for the vascular dementia outcome, MR studies of this condition are precluded). This will restrict the triangulation exercise for this outcome, as it will have to rely on different analytical strategies within a single study design (non-randomised studies) rather than incorporating fundamentally different study designs.

A final point on the progress of the review is that, due to the need for dual screening and data extraction, a number of external researchers became involved in this review. I found the people-management aspect particularly challenging and could definitely have improved the process through better communication of deadlines, but it has provided good experience of leading a review team.

**Added value**

One of the take-aways from my last progress review was to make the “added value” of this work beyond existing reviews clear. I believe there are three aspects where this review is distinct from those reviews already available in the published literature:

* *Comprehensiveness:* While several reviews of this research topic exist,1–4 the overlap between the list of studies included in each is not 100%. As part of this review, I have not only performed a original search of primary literature databases, but have also screened the reference lists of comparable reviews to ensure no study has been omitted.
* *Structured risk of bias assessment:* The majority of the highly cited reviews on this topic either do not formally consider the risk of bias in the observational studies they include or do not use an appropriate domain-based assessment tool (e.g.ROBINS-I/E). This is important area in which this thesis can add value, as based on the risk-of-bias assessments I have performed to date, several primary studies are at high risk of bias and this should be reflected in the findings of any review on this topic.
* *Inclusion of preprints:* Unlike other available reviews and enabled by the tool described in Chapter 3, this review systematically searched preprinted health-related manuscripts as a source of grey literature. As part of this chapter, I plan to examine the extent of the additional information provided to the review by the inclusion of preprints.

## Chapter 5: Observational analysis

**Overview**

This chapter uses data from the Clinical Practice Research Datalink (CPRD) to investigate the relationship between statin use and dementia.

Rather than reproduce the methods/findings of this chapter here, I have attached the initial draft of the paper this chapter is based on at the end of this report (Section 6).

**Progress**

* Paper: 80%
* Chapter: 80%

In relation to the results presented in the attached draft, there are two key discussion points. The first is that while statins are associated with a slight reduction in the risk of probable (HR: 0.91, 95%CI: 0.88-0.95) and possible Alzheimer’s Disease (HR: 0.95, 95%CI: 0.92-0.99), there is an increased risk of vascular dementia (HR: 1.7, 95%CI: 1.62-1.77) and other dementia (HR: 1.13, 95%CI: 1.09-1.18) in those using statins. This reflects the initial findings of this analysis presented at the last progress review, and was not attenuated even after using multiple imputation to account for potential selection bias caused by the original complete case analysis and after accounting for several potentially important confounders.

Secondly, following circulation of the draft, a co-author reccommended that I use the same analysis strategy with a known control (coronary heart disease; CHD). However, when running this analysis, I found that statins appeared to increase the risk of CHD, in contrast to the the results of published large RCTs. This finding suggests the presence of substantial confounding by indication in this analysis, in that those who go on to take a statin are fundamentally different from those who do not, and this has been seen in other standard mulitvariable HR analyses of the effect of statins on CHD.5 A suggested approach to frame this (possible insurmountable) issue has been to rewrite the paper as an example of the dangers of using EHR in cases where strong confounding by indication is likely. However, it would be particularly useful to get reviewers feedback on any further analytical strategies Icould attempt to address this issue.

In terms of the broader thesis, while these concerns means I have low confidence in the accuracy of the result, it presents a good opportunity for comparison with other published papers that used different analytical strategies where the source and direction of bias is different (e.g. confounding by indication in our analysis likely makes statin use look worse but immortal time bias in other papers likely makes statin use look better).

Finally, in an attempt to assess whether these unexpected findings for vascular dementia are a product of the CPRD data or our analytical strategy, I recently attempted to replicate an analysis performed by Smeeth et al. which used the THIN EHR database and found a protective effect of statins on dementia risk (HR 0.81, 99% CI 0.69-0.96).6 However, following correspondence with the author team to acquire the code-list they used to define dementia cases, I confirmed that it was substantially different from that used in our study. This fact, coupled with concerns over their analytical strategy (which seemed to adjust for covariates on the causal pathway), limited the value of this attempted replication.

**Added value**

Similar to the systematic review presented in Chapter 4, feedback from my last progress review suggested that the “added value” of this study above those that already exist should be made clearer. There are two particular motivations for this analysis above those already available in the published literature:

* *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:* Analysing 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.7 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.

## Chapter 6: Individual patient data meta-analysis

**Overview**

This chapter will use data from several Dementia Platform cohorts (plus some of the cohorts identified through the systematic review) to investigate the relationship between blood lipids levels and dementia outcomes. This analysis has not yet started, but data access has been secured.

**Progress**

This analysis has not yet formally started, though access to several DPUK cohorts has been secured. Cohorts identified through the systematic review have also been invited to participate (though I am sceptical about the chance of response).8

## Chapter 7: Triangulation

**Overview**

This chapter will draw together different sources of information (including the primary analysis performed as part of this thesis), along with a consideration of the key sources/direction of bias in each, in a qualitative triangulation framework.9 This will ve

As flagged in my last progress review, this chapter will make very clear that “triangulation” in the context of my thesis is a qualitative integration of the distinct evidence sources available (systematic review, CPRD study, IPD meta-analysis).

**Progress**

Not started.

## Chapter 8: Discussion/limitations/conclusion

**Overview**

This chapter a summary of the main findings of the thesis, a discussion of the strengths and limitations of the work presented, a roadmap for future work, and a conclusion.

**Progress**

Not started.

# Other

## Courses

All training courses planned for the past year have either been cancelled or postponed due to the pandemic. As a result, in the review period, I have not taken part in any courses related to my thesis.

However, in the coming year, I plan to attend the “Mendelian randomisation” and “Advanced methods for multiple imputation” short courses at the University of Bristol, in addition to the “Causal Inference in Epidemiology: Recent Methodological Developments” short course at LSHTM and the “Writing about your research” course at the Royal Society.

## Conferences/ presentations

In the review period, I presented on the preprint search tool described in Chapter 3 at the Bristol “Autosynthesis Club” and “Methods in Evidence Synthesis” seminars.

It’s only in writing this report that I realised that I am yet to present on the other parts of my thesis, and so this is a particular aspect of my PhD that I would like to develop further in the coming year. As such, I aim to present at more diverse events in the coming year, including at an IEU monthly meeting and the Faculty of Health Sciences research showcase.

In addition, I plan to submit an abstract to the Alzheimer’s Association International Conference in July (deadline 25th January).

## Teaching

In the review period, I have gained the following teaching experience:

* MSc Epidemiology/Public Health “Clinical Epidemiology” module
  + Tutored a small group of several weeks
  + Marked the end-of-module assessments
* “Introduction to R” short course for new PhD students, Bristol Medical School short course
  + Presented a lecture on advanced R topics (data visualisation, literate programming and web applications)
* “Introduction to Data Visualisation and Web Applications using R”, Bristol Medical School short course
  + Designed and delivered the web applications aspects of the course
  + Tutor helping on practicals on data visualisations and literate programming with Rmarkdown
* Drop-in “Intro to R” sessions
  + Organised and ran an hour long session each Friday in the autumn term to help answer new PGR students’ questions about R.

I have also submitted a proposal to supervise an MSc student for the summer term. The project aims to examine the overlap between the set of primary studies included in the several existing systematic reviews of blood lipid levels and dementia (as mentioned in the description of Chapter 4, above).

# Other work

During my PhD I have been involved in a number of project unrelated to my thesis.

**COVID-related work**

* *Living review:* In early March, I developed and deployed an integrated literature searching and screening pipeline to help a team of national experts produce the first iteration of a living systematic review on the impact of COVID-19 on suicide and self-harm (now published: <https://doi.org/10.12688/f1000research.24274.1>). Maintenance of this pipeline took up a lot of my time - however, following discussions with my primary supervisory, in late summer, I handed over responsibility to another member of the team.
* *Rapid review of symptoms:* Through my work on medrxivr (the preprint search tool described in Chapter 2), I was involved in developing and running preprint searches for a rapid systematic review of COVID-19 symptoms (now published: <https://dx.doi.org/10.2139/ssrn.3582819>)

**Evidence synthesis methods**

* *robvis:* A paper describing the risk-of-bias visualisation tool I developed was published as part of special issue of “Research Synthesis Methods” (<https://doi.org/10.1002/jrsm.1411>). The tool has now been used ~100 published reviews.
* *PRISMA2020:* I contributed to the updated PRISMA2020 guidelines on the preferred reporting items for systematic reviews and meta-analyses, and am a co-author on two papers describing the new version of this reporting checklist which have been accepted for publication in the BMJ.

**Primary research**

* *Comparison of data availability statements:* Using medrxivr (the preprint search tool described in Chapter 2), myself and a fellow PhD student compared the availability of data for the same manuscript when published under two different data-sharing policies (i.e. that of the preprint repository vs. the peer-reviewed journal). The preprint of the study is available [here](https://doi.org/10.31222/osf.io/p75xe).
* *Making primary research synthesis-ready:* I acted as last author on a commentary describing the best practice and advatnages for primary researchers in the field of prevention science to ensure that their research is “synthesis-ready” (i.e. can be easily found and integrated into an evidence synthesis project). This commentary is under consideration at “Prevention Science” and a preprint is available [here](https://doi.org/10.31222/osf.io/ptg9j).

## Report references

1. 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).

2. Yang, Z. *et al.* Association of blood lipids, atherosclerosis and statin use with dementia and cognitive impairment after stroke: A systematic review and meta-analysis. *Ageing Research Reviews* **57**, 100962 (2020).

3. Muangpaisan, W., Brayne, C. & the Alzheimer’s Society Vascular Dementia Systematic Review Group. Systematic review of statins for the prevention of vascular dementia or dementia. *Geriatrics & Gerontology International* (2010). doi:[10.1111/j.1447-0594.2009.00579.x](https://doi.org/10.1111/j.1447-0594.2009.00579.x)

4. 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).

5. Danaei, G., Rodríguez, L. A. G., Cantero, O. F., Logan, R. & Hernán, M. A. Observational data for comparative effectiveness research: An emulation of randomised trials of statins and primary prevention of coronary heart disease. *Statistical Methods in Medical Research* **22**, 70–96 (2013).

6. Smeeth, L., Douglas, I., Hall, A. J., Hubbard, R. & Evans, S. Effect of statins on a wide range of health outcomes: A cohort study validated by comparison with randomized trials. *British Journal of Clinical Pharmacology* **67**, 99–109 (2009).

7. 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).

8. Nevitt, S. J. *et al.* Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: Systematic review. *BMJ* **357**, j1390 (2017).

9. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology. *International Journal of Epidemiology* **45**, 1866–1886 (2016).

# Draft of CPRD paper

**Title:**

Association of lipid regulating drugs with dementia and related conditions: an observational study of data from the CPRD

**Authors and Affiliations:**

Luke A McGuinness1,2, Venexia M Walker1,2, Neil M Davies1,2, Richard M Martin1,2, Yoav Ben-Shlomo2, Julian PT Higgins1,2, Elizabeth Coulthard^\*\*\*^, George Davey-Smith1,2 and Patrick G Kehoe3,4

* MRC University of Bristol Integrative Epidemiology Unit, Bristol, UK
* Bristol Medical School: Population Health Sciences, University of Bristol, Bristol, UK
* Dementia Research Group, Bristol Medical School, University of Bristol, Bristol, UK
* Bristol Medical School: Translational Health Sciences, University of Bristol, Bristol, UK

**Corresponding author:**

Luke McGuinness; Bristol Medical School, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, United Kingdom.; [luke.mcguinness@bristol.ac.uk](mailto:luke.mcguinness@bristol.ac.uk)

##### 

# Introduction

Dementia is a major neurocognitive disorder, the most common types of which are Alzheimer’s disease (AD) and vascular dementia (VaD).1 Despite an increasing number of cases globally and decades of research, there remains much unknown about the pathogenesis and progression of the disease, and, at present, no effective treatment exists to arrest or reverse the cognitive decline associated with the condition.2 Drug repurposing, the identification of new applications for previously approved drugs the may provide an efficient mechanism by which to discover effective preventative and therapeutic treatments for dementia.3,4

Several cardiovascular elements have been identified as potential risk factors for dementia,5 and of these, lipid levels represent a promising target for intervention due to the ready availability of lipid-modifying treatments. In this context, determining whether lipid regulating agents (LRAs) could be repurposed for the prevention of dementia and related diseases would be helpful in the development of evidence-based prevention policy. Several existing prospective studies have examined the association of lipid regulating agent use with dementia exist.6–8 However, many of studies are small, record few outcomes, and have limited follow-up.

The use of such electronic health data for epidemiological research has several advantages.9 As the data are collected through the routine care of a large cohort, they allow for retrospective and prospective studies using sample sizes and time-scales which would be unfeasible using traditional methods. In addition, data are collected for care provision and without a specific research question in mind, providing a holistic picture of a patient and their health experience. This provides vital data on a range of potential confounders which can be incorporated into an analysis.

We therefore aim to examine the association between several major classes of LRAs and all-cause dementia, Alzheimer’s disease, vascular dementia and other dementia, in the Clinical Practice Research Datalink (CPRD), a large, population-based electronic health record (EHR) database.

##### 

# Methods

## Study design and protocol

We performed a prospective cohort study using data from the CPRD (Clinical Practice Research Datalink), a large UK primary healthcare database.10 Our initial sample included all participants included all participants registered at a participating practice between 1-1-1995 and 29-2-2016 who had a flag for “research quality” data. Index, exposure and outcome events were identified using predetermined code lists, which are available for inspection (see [Data/code availability](file:///H:\SafeHaven\CPRD%20Projects%20UOB\Projects\15_246\CPRD-LRA\manuscript\CPRD_LRA_Draft.docx#data-code-avail)).

An *a priori* protocol for this study was published,11 and amendments to this are recorded in Supplementary Table 1. This study was reported in line with the STROBE Cohort guidelines (Supplementary Table 3).12

## Study Cohort

Participants were included in our study cohort if their record contained any of the following index events: a READ code code for a diagnosis of hypercholesterolemia or related condition; a READ code for prescription of a lipid regulating agent (statin, fibrates, ezetimibe, etc.); a total cholesterol test result of >4mmol/L; or a 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/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 address confounding by indication, as participants with similar index events are either subsequently prescribed or not prescribed to a drug. If the total cohort was included instead, the unexposed group would likely be healthier across a range of variables than those in the exposed group, leading to a biased association been lipid regulating agent use and dementia. Conditioning entry on being either “at-risk” or already diagnosed with hypercholesterolemia attempts to mitigate this bias.

The index date was defined as the date where the first relevant code/test results was recorded on their record, and participants were followed up until the earliest of: an outcome of interest; death; end of follow-up (29-2-2016); or last registration with their GP practice. Participants were removed from our sample if they had less than 40 years of age, had less than 12 months of “research quality” data, were initially prescribed more than one lipid regulating agent, or were diagnosed with dementia before/on the date of the index event.

## Exposures

We considered 7 lipid regulating drug classes based on groupings in the British National Formulary (BNF)13, namely: statins, fibrates, bile acid sequestrants, ezetimibe, nicotinic acid groups, ezetimibe and statins, and omega-3 fatty acid groups.

A participants drug class was assigned based on their first recorded prescription, and any drug switching was not ignored in an effort to mimic an intention-to-treat approach. We did however examine the frequency of drug class stopping (the last prescription of the primary class was followed by at least 183 days of observation), adding (a second drug class was prescribed before the last prescription of the primary class), and switching (a second class was prescribed after the last prescription of the primary 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. When two or more outcomes were recorded in a participants record, a decision tree was used to differentiate between them (Supplementary Figure 1). The diagnosis date of the neurodegenerative disease was determined by the first record of a relevant code.

## Analysis plan

All analyses were performed in STATA 15. Cox proportional hazard models were used to estimate the hazard ratio and corresponding 95% confidence intervals. To address the potential for immortal time bias, we employed a time-varying indicator of treatment status in order to correctly allocate time-at-risk to the exposed and unexposed groups. The time axis for the model was participant age.14–16 The final model was adjusted for a range of baseline covariates including sex, 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, or hypertension.

In the case of missing data, we used a multiple imputation by chained equations (MICE) approach in STATA to create 10 imputed datasets. All covariates included in the analytic model were also included in the imputation model. The full imputation model is available for inspection (See [Data/Code availability](file:///H:\SafeHaven\CPRD%20Projects%20UOB\Projects\15_246\CPRD-LRA\manuscript\CPRD_LRA_Draft.docx#data-code-avail) section).

Analyses were performed by outcome and by drug class. For the drug class analyses, those prescribed a different drug than the class under investigation were excluded. Similarly, for the dementia sub-types analyses, participants with an alternative diagnosis were excluded.

## Sensitivity analyses

In an effort to address the potential for reverse causation (where pre-clinical pathology could influence factors that prompt the prescription of lipid regulating agents), we introduced an increasing delay between initial exposure and diagnosis. We excluded events occurring within an increasing delay window but kept the time-at-risk, so as to avoid re-introducing immortal time bias. We used this delay as a proxy for extended exposure to vascular risk factors that prompt LRA use.

We also stratified by grouped year of entry into the cohort (<2000, 2000-2004, 2005-2009, >2010) to explore the potential for a time period effect.

**Suggestions for further sensitivity analyses welcome!**

## Data/code availability

We used the CPRD-GOLD primary care dataset March 2016 snapshot (ISAC 15\_246R), which is available upon application to the CPRD Independent Scientific Advisory Committee. The code lists used to define the outcomes and covariates for this study, in addition to the cleaning and analysis scripts used to create the study cohort and perform the analyses, are available on Github (<https://github.com/mcguinlu/CPRD-LRA>).

##### 

# Results

## Patient characteristics

Of the 3179733 participants included in our extract, 1711381 met the inclusion criteria (See Supplementary Table 2 for the attrition flowchart), with a total follow-up of 11,091,437 patient years at risk. The median age at index was 57 years (IQR:48-67) and participants were followed up for a median of 5.9 (IQR:2.7-9.7). During follow-up, an all-cause dementia diagnosis was recorded for 42313 patients (12751 probable AD, 10107 possible AD, 8542 vascular dementia, 10913 other dementia). The distribution of baseline characteristics across the drug classes can be seen in Table 1.

Table 1: Patient characteristics by drug class

|  | **Whole Sample** | **No LRA** | **Statins** | **Bile acid sequestrants** | **Ezetimibe** | **Ezetimibe & Statins** | **Fibrates** | **Nicotinic acid groups** | **Omega-3 Fatty Acid Groups** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **N** | 1711381 | 1118651 | 581499 | 5350 | 753 | 126 | 3851 | 164 | 987 |
| **Median year of index event/prescription** | 2006 | 2007 | 2004 | 2005 | 2004 | 2005 | 2001 | 2001 | 2005 |
| **Female** | 53.1% (908324) | 56.1% (627791) | 47.2% (274402) | 66.5% (3560) | 54.4% (410) | 53.2% (67) | 38.6% (1486) | 54.9% (90) | 52.5% (518) |
| **Age** | 57 | 54 | 62 | 57 | 60 | 57 | 58 | 62 | 56 |
| **CAD** | 0.4% (6956) | 0.1% (640) | 1.1% (6238) | 0.1% (6) | 0.9% (7) | 0.0% (0) | 1.4% (52) | 0.0% (0) | 1.3% (13) |
| **CBS** | 0.3% (5636) | 0.1% (737) | 0.8% (4810) | 0.1% (4) | 0.3% (2) | 0.0% (0) | 2.0% (77) | 0.0% (0) | 0.6% (6) |
| **CVD** | 2.1% (35196) | 1.1% (12260) | 3.9% (22635) | 1.6% (86) | 2.5% (19) | 2.4% (3) | 4.4% (169) | 4.3% (7) | 1.7% (17) |
| **Charlson (ever > 0)** | 30.5% (522725) | 25.3% (283183) | 40.3% (234530) | 42.2% (2258) | 41.3% (311) | 24.6% (31) | 50.5% (1945) | 43.3% (71) | 40.1% (396) |
| **IMD-2010 (median)** | 9 | 8 | 9 | 8 | 9 | 13 | 10 | 10 | 10 |
| **Consulation rate (mean/SD)** | 5.4 (5.4) | 5.0 (5.0) | 6.2 (6.0) | 8.5 (7.3) | 7.3 (6.6) | 4.8 (4.3) | 7.0 (6.0) | 9.2 (7.8) | 8.0 (8.0) |
| **Alcohol (ever)** | 85.9% (1469813) | 86.5% (967461) | 84.8% (493036) | 82.9% (4436) | 84.2% (634) | 88.1% (111) | 82.8% (3188) | 83.5% (137) | 82.1% (810) |
| **Smoking (ever)** | 51.1% (873795) | 47.1% (526337) | 58.7% (341051) | 55.3% (2960) | 58.0% (437) | 59.5% (75) | 60.3% (2322) | 52.4% (86) | 53.4% (527) |
| **BMI (mean/SD)** | 27.0 (5.3) | 26.7 (5.2) | 27.7 (5.2) | 26.8 (5.8) | 28.0 (5.7) | 28.2 (4.8) | 28.9 (5.2) | 26.5 (5.0) | 26.8 (5.4) |
| **PAD** | 0.7% (12622) | 0.4% (4234) | 1.4% (8241) | 0.9% (46) | 0.8% (6) | 0.8% (1) | 1.9% (74) | 6.1% (10) | 1.0% (10) |
| **Hypertension** | 15.9% (272095) | 11.6% (129624) | 24.1% (140431) | 12.5% (667) | 22.8% (172) | 25.4% (32) | 25.5% (981) | 21.3% (35) | 15.5% (153) |
| **Total cholesterol (mean/SD)** | 5.8 (10.1) | 5.5 (6.3) | 6.3 (15.5) | 5.6 (1.0) | 7.3 (27.0) | 6.8 (1.4) | 6.6 (5.7) | 5.6 (1.1) | 5.8 (1.4) |
| N - Numer of included participants; IMD - Index of Multiple Deprivation;BMI - Body Mass Index; CAD - Coronary Arterial Disease; CBS - Coronary Bypass Surgery; CVD - Cardiovascular disease; PAD - Peripheral arterial disease. | | | | | | | | | |

A substantial majority (98.11%) 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 1).

## Missing data

Full covariate information was available for 773305 participants (45.19%). Five 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 636394 participants (37.19%); alcohol status was missing for 273918 participants (16.01%); smoking status was missing for 85862 participants (5.02%); BMI, or a calculated BMI from height and weight measurements, was missing for 270923 participants (15.83%), and baseline total cholesterol was missing for 148588 participants (8.68%).

## Primary analysis

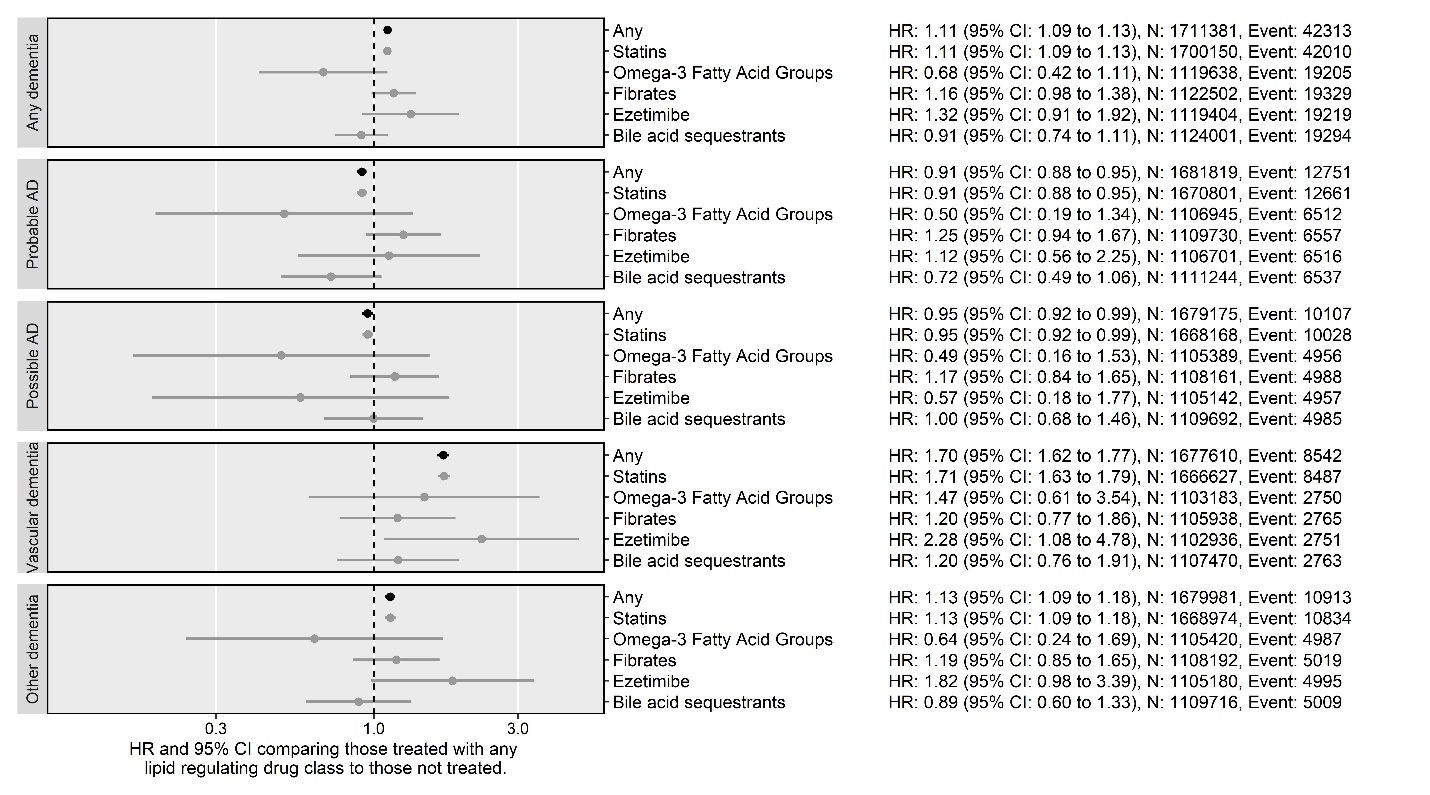


Figure 1: Results from the Cox proportional hazards analysis, using age as the time scale, by outcome and drug subgroup. AD: Alzheimer’s disease

**Alzheimer’s disease**

As shown in Figure 1, our results suggest a small protective effect of lipid regulating agents on probable (HR: 0.91, 95%CI: 0.88-0.95) and possible (HR: 0.95, 95%CI: 0.92-0.99) Alzheimer’s disease when compared to no treatment.

**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: 1.7, 95%CI: 1.62-1.77) or other dementia (HR: 1.7, 95%CI: 1.62-1.77). Again this effect was driven mainly by the statin subgroup, but of note, there was some evidence that ezetimibe was associated with an increased risk of vascular dementia (HR: 2.28, 95%CI: 1.08-4.78).

**All-cause dementia**

For the composite all-cause dementia outcome, we found treatment with a lipid regulating agent was associated with a slightly increased risk (HR: 1.11, 95%CI: 1.09-1.13), but the magnitude of the association was not as extreme as that observed for the vascular and other dementia subgroups.

## Sensitivity analyses

Supplementary Figure 3/4 show the results of our secondary analyses. When incorporating an increasing delay between start of exposure and diagnosis, we observed a large effect estimate at longer delays, indicating that duration of exposure results has some effect on subsequent risk of dementia and related conditions (Supplementary Figure 3).

For the time period analysis, we observed no variation in risk by time period in any subgroup except for probable Alzheimer’s disease (Supplementary Figure 4).

**Other secondary/sensitivity analyses TBC**

##### 

# Discussion

## Main findings

Lipid regulating agents reduced the risk of probable and possible Alzheimer’s when compared with no treatment, but increased the risk of an all-cause dementia, vascular dementia and other dementia diagnoses. 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 sole exception being that ezetimibe appears to increase the risk of vascular dementia.

## Comparison to other literature

Much of the existing literature is focused on the the association of statins alone with neurodegenerative outcomes, with other lipid regulating agents being grouped as “non-statin cholesterol-lowering drugs”.17 This echos 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 vs non-treatment for the prevention of the dementia, only one of which presented information on the incidence of dementia.18 This study (Heart Protection Study) showed no effect of treatment with pravastatin on all-cause dementia risk (OR: 1.00, 95%CI:0.61-1.65),19 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 0.832, 95%CI: 0.793–0.872).

Both of these alternative sources of evidence conflict with the findings of our analysis, where statin use was associated with an increased risk of all-cause dementia (HR: 1.11, 95%CI: 1.09-1.13). However, some of the included studies in the meta-analysis specifically exclude vascular dementia from the definition of all-cause dementia,20 while others, including a large scale study of the QResearch database, another large scale England-based EHR database, are likely at risk of immortal time bias, having followed statin users and non-users from different index dates.21,22

**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.23 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).24

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

Significantly 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).23 This contrasts with the increased effect found in our analysis (HR: 1.7, 95%CI: 1.62-1.77) (need to flesh this out).

**Other drug classes**

For the drug class analysis, our findings agreed with the limited available literature, as 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,17 in agreement with our findings.

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 vascular dementia.

## 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, unlike some previous large scale EHR analyses,22 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. In this case, the all-cause dementia outcome could be seen as a quasi-sensitivity analysis to assess the impact of this misclassification, and does seem to fall between the Alzheimer’s disease and vascular/other dementia results.

We attempted to address confounding by indication, where factors that affect whether a participants is exposed also affect their outcome, by limiting inclusion to those either prescribed or “at risk” of being prescribed, based on an initial indicative elevated test result or diagnosis, a lipid regulating agent. We also adjusted for several additional potential confounding variables. However, the analysis could still be biased by residual confounding, including by 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.25 This finding, combined with the potential for misclassification between Alzheimer’s disease and vascular dementia, could go someway to explaining the observed association of lipid regulating agents with increased 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. We have attempted to address this by introducing a delay in a secondary sensitivity analysis.

##### 

# 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 lipid regulating agents had a small protective effect against probable and possible Alzheimer’s disease, but were associated with an increased risk of all-cause, vascular and other dementia. In all cases, the estimated effect was driven by that observed in the statin sub-group, which made up 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.