# 1 Discussion

## 1.1 Lay summary

**TBC**

## 1.2 Introduction

The central aim of this thesis was to infer the causal effect of blood lipid levels on dementia outcomes via evidence synthesis methods. In this discussion, I will review the contribution of each chapter to this aim, summarise the principal findings and methodological innovations of the thesis, and discuss the implications of my findings for both clinical and public health practice. I consider the overall strengths and limitations of this thesis as a coherent body of work and suggest avenues for impactful future research that builds on the research presented.

### 1.2.1 Chapter summary

This thesis is comprised of a mix of methodological and applied research chapters.

In **Chapter ??**, I introduced the core concepts considered in this thesis and framed the research presented in relation to a theoretical framework of evidence synthesis.

In **Chapter ??**, I presented a new tool for the systematic searching of health-related preprints, and detailed two use cases for preprint metadata beyond the systematic searching of this evidence source.

In **Chapter ??**, I described the methods of a comprehensive systematic review of all available evidence on the association of both lipids and lipid-regulating agents with dementia outcomes. The results of this analysis are then presented in **Chapter ??**.

In **Chapter ??**, I described an analysis of the association of lipid-regulating agent use and dementia outcomes in the CPRD, a large electronic health record database. I make use of control outcomes to illustrate insufficiently controlled confounding.

In **Chapter ??**, I presented an individual participant data meta-analysis of the association of lipid levels and dementia outcomes, using previously unanalysed data accessed via the Dementia Platform UK.

Finally, in **Chapter ??**, I proposed a new method for the systematic integration of multiple evidence sources as part of quantitative triangulation framework. I illustrated the method using two case studies (the effect of LDL-c on Alzheimer’s disease and the effect of triglycerides on vascular dementia), drawing on the evidence identified or produced by previous chapters.

## 1.3 Summary of clinical findings

### 1.3.1 Effect of blood lipids on dementia outcomes

As the results of the different components of this thesis are discussed in detail as part of the triangulation analysis (see Section ??), here I briefly summarise the key findings.

Overall, I did not identify a consistent effect of any blood lipid fraction (total cholesterol, LDL cholesterol, HDL cholesterol or triglycerides) on any dementia outcome (all-cause dementia, Alzheimer’s disease or vascular dementia). While published non-randomised studies provided evidence for a protective effect of lipid-lowering via statin use on Alzheimer’s disease and the IPD analysis suggested a harmful effect of raised triglycerides on vascular dementia, these effects were not maintained when incorporated along with other sources of evidence in a quantitative triangulation framework. Similarly, the IPD analysis did not provide evidence for an interaction of patient characteristics with the association of blood lipids on dementia outcomes, though the results are limited by low number of cohorts analysed.

In the published literature, statins were by far the most studied lipid-regulating agent in relation to dementia outcomes. Given the large proportion of patients taking statins as indicated by my analysis of the CPRD data, this finding is unsurprising. Finally, there was a substantial absence of evidence for vascular dementia.

### 1.3.2 Comparison with new evidence

Since the searches underpinning the systematic review described in this thesis were performed (July 2019), further studies on this topic have been performed. Two of these, notable for the analysis of large UK cohorts, are discussed here.

The first is a large-scale analysis of 953,635 patients in the CPRD, followed from their first LDL-c measurement.1 The study found a slight increased risk of dementia was associated higher LDL-c measured at mid-life, driven by the finding for the Alzheimer’s disease subgroup. These results are consistent with the findings presented in this thesis, particularly given the lack of adjustment for confounding by ApoE4, which was estimated by the study to induce a RR of 1.09 per 1-SD increase in LDL-c in the absence of a true causal effect.

The second analysis examined 502,226 participants in the UK Biobank, a large population-based prospective cohort study.2,3 The analysis found no association between lipids measured at mid-life and dementia outcomes, comparable to the results of this thesis. However, similar to the study discussed above, this analysis did not have access to genetic data and so may be subject to residual confounding by *ApoE4* genotype.

### 1.3.3 Implications for clinical practice

Given the absence of any consistent signals across the evidence sources analysed, there are no clear indications for clinical practice arising from this thesis. While my research did not provide any strong evidence for an effect of lipid-lowering on dementia outcomes, the cumulative strength of evidence is also insufficient to rule out an effect.

Given this ambiguity, the use of statins should be restricted to their primary and well-understood purpose of lipid-lowering to reduce the risk of cardiovascular events. However, it is important to recognise that even under the scenario that a strong harmful effect of lipid-lowering on dementia outcomes was identified, whether or not to prescribe a statin at mid-life will always be guided by factors other than dementia prevention. Patients must live to a certain age in order to be at risk of dementia (with the exception of familial early-onset dementias), which is substantially less likely if hypercholesterolemia at mid-life is left unchecked.

In summary, clinicians should be aware of the uncertainty of evidence surrounding the effect of lipid-lowering on dementia outcomes and should be prepared to convey the same to patients.

### 1.3.4 Implications for public health

**[Note: Yoav, I am particularly interested in your thoughts on this section]**

Similarly, given the ambiguity of evidence, there are no clear implications for public health practice. However, the role of public health, and particularly continuous population surveillance, may be important in providing a further source of evidence in relation to the effect of lipid lowering on dementia. Public health practitioners could examine summary-level statistics of GP prescription data to attempt to relate lipid-lowering treatment to temporal or geographical trends in dementia incidence. It will be difficult to assess if any observed trends are due to a true effect or to confounding or ecological biases. However, as dementia outcomes will be monitored anyway for incidence/prevalence estimation and planning of services/care, this approach would provide a ready source of further evidence. For example the CFASII study showed a reduction of age-specific prevalence rates compared to CFASI which may be related to better blood pressure control. (Ref Matthews Lancet) The use of natural experiments, e.g. prescription rates of statins both within and between countries, could be used to test the statin hypothesis, though potential cohort effects on dementia need to be considered.

## 1.4 Summary of methodological contributions

In addition to addressing the causal impact of lipids on dementia outcomes, the research described in this thesis represents a number of novel contributions to the field of evidence synthesis methodology.

### 1.4.1 Inclusion of preprinted evidence

Preprints are an important source of evidence and searching of preprint repositories should become an accepted part of the evidence synthesis process. To support this, I developed a new tool (as described in Chapter ??) to enable researchers to readily search health-related preprints.

### 1.4.2 Bias

One of the central methodological contributions of this thesis has been on the conduct, visualisation, and incorporation of bias assessments as part of an evidence synthesis exercise.

In terms of assessment, I applied domain-based tools to evidence relevant to the effect of lipids on dementia outcomes and highlighted the limitation of the available tools for Mendelian randomisation studies. I also piloted and provided feedback on a forthcoming tool for assessing the risk of bias due to missing evidence. In the primary analysis of data from the CPRD, I used negative control outcomes to assess the potential for bias due to insufficiently controlled confounding by indication.

In terms of visualisation of bias, I have made a number of methodological contributions. I created a well-received R package to create paired forest plots (see Appendix ??), and built on this to enable creation of the bias direction plots illustrated in Chapter ??.

Finally, I made several novel contributions around the incorporation of bias into analyses. These were intended to encourage authors of reviews to actively consider the impact of bias in their synthesis, something that empirical evidence suggests happens infrequently at present.4 The new methods introduced by this thesis range from the forest plots stratified by overall risk of bias level presented in Chapter ?? to the more advanced triangulation framework presented in Chapter ??.

### 1.4.3 Triangulation framework

The proposed quantitative triangulation framework represents the final novel contribution of this thesis to evidence synthesis methodology. Using pre-specified adjustment distributions mapped to the results of the domain-based risk-of-bias and indirectness assessments, this approach provides a systematic way to account for biases/indirectness across studies, reducing the potential for differential adjustment based on knowledge of the results.

Ideally, this framework should be based on the results of a comprehensive systematic review, which identifies all evidence (both direct and indirect) related to a causal question of interest. This would represent a break from the current practice of limiting review to a single type of evidence, and would result in fewer, more comprehensive reviews. However, in the current context of research waste, a concentration of effort in fewer reviews may be advantageous. Section ?? illustrates the duplication of reviews in relation to the topic of this thesis, while an assessment of the COVID-19 literature found that, for a specific clinical question, there were substantially more reviews (25) than available primary studies (17).5

Finally, while the framework proposed here could (fairly) be criticised over the validity of the prior distributions of bias/indirectness chosen, the assumptions made about the impact of bias in this thesis are no stronger than those made when synthesising effect estimates as though they were unbiased. The proposed framework at the very least recognises the uncertainty introduced by bias/indirectness and decreases the precision of the overall effect estimate on this basis.

In summary, the proposed triangulation framework will enable better use of all available evidence to address causal questions.

## 1.5 Overall strengths and limitations

The strengths and limitations of each aspect of this thesis have been discussed in the respective chapter. Here, I highlight the strengths and weaknesses of this thesis as a body of work.

### 1.5.1 Strengths

There are several strengths to this thesis as a whole. Specifically, the identification and triangulation of evidence across study designs and publication status sets this thesis apart from previous synthesis of the evidence on this topic. Additionally, this thesis has also produced new evidence on the association of blood lipids with dementia outcomes in previously unanalysed cohorts and provided additional information on a previously under-studied association (lipids/LRA and vascular dementia).

A further strength of this thesis is the production of software to support the novel evidence synthesis techniques used. Extensive documentation has been written to guide users through usage of the tools and example data is provided for use. Finally, substantial effort has been invested to make the research documented in this thesis as reproducible as possible - this is documented more fully in Appendix ??.

### 1.5.2 Weaknesses

However, there are also several strong limitations to this thesis. In the first instance, as with many studies of dementia outcomes, a limiting factor in the interpretation of the results is the absence of a detailed pathological mechanism. It is possible that statins have a true effect via some other pathway , but without knowledge of the mechanism, the plausibility of this relationship cannot be assessed.

Secondly, each analysis presented here makes use of secondary data sources, be it published literature, electronic health records, or existing cohorts. Secondary data can limit the type of analyses performed (for example, in the IPD analysis presented in Chapter ??, absence of time-to-event data prevented the use of hazard ratios to quantify dementia risk) and the validity of the data if collected for purposes other than research (for example, the accuracy of dementia diagnoses in electronic health records is known to be variable).6

Thirdly, missing evidence was a common limitation across this thesis. In Chapters ??/??, several studies stated that they did not report the results of an analysis because the results were not significant. This was particularly common among non-randomised studies of exposure examining blood lipids directly. Similarly, an absence of evidence on vascular dementia, potentially due to a publication bias mechanism, limited the analysis of this outcome. Missing evidence was also as an issue in the IPD meta-analysis reported in Chapter ??, in the form of a poor response to data access requests, while the absence of empirically-based distributions of bias/indirectness for use with the triangulation framework limited the credibility of the results produced.

A final weakness stems from the geographical focus of the data analysed in this thesis. All primary analysis presented drew on data from the UK (CPRD, CaPS, EPIC Norfolk, Whitehall II), while the majority of studies identified by the review were based in the Western world (Figure ??). This may limit the generalisability of the results presented to different populations.

## 1.6 Lessons learned

As part of a reflective learning process throughout my PhD, I maintained a catalogue of failures (available in Appendix ??) describing analytical mistakes, failed experiments, and unsuccessful grant applications. However, I found there was one central learning point from this thesis, namely to be slightly less ambitious when planning future research projects.

The proposal for the program of research underpinning this thesis was created in advance of starting my PhD, as a detailed plan was required in order to secure the funding that supported me through my studies. In hindsight, the decision to take a broad approach to the inclusion of dementia outcomes resulted in a larger workload than anticipated, particularly when also considering evidence from multiple different study designs. Additionally, as dementia subtypes likely have very different aetiological pathways, it may have been better to focus on a single subtype (e.g. Alzheimer’s disease) when considering the causal effects of lipids. Similarly, attempting to undertake a full individual participant data meta-analysis as a single part of a larger program of research was overly ambitious, as data cleaning and harmonisation for just three cohorts was a substantial undertaking.

In the future, armed with the experience gained through conduct of the research projects presented here, I will be better place to scope and design research projects.

## 1.7 Future work

Several avenues of future research could be pursued, building on the novel work presented here. These are grouped by topic in the following sections.

**[Question: is it better to use “should” or “could” when listing ideas for future research? I.e. “future work could/should look at X”?]**

### 1.7.1 Generation of new RCT evidence

It is customary at this point in the discussion of results based (primarily) on observational data to recommend that a large-scale randomised controlled trial be performed to assess the effect of the proposed intervention. Given the costs and logistical challenges associated with trials examining outcomes with long prodomal periods such as dementia, a more efficient approach is offered by the opportunistic post-trial follow-up (PTFU) of existing RCTs of lipid-regulating agents.

This approach has already been used to assess long-term (11-20 years) safety outcomes of statin use, using participant recontact7 and linkage with electronic health records to identify events.8 A similar approach could be used to assess the impact of randomisation to statins at midlife on subsequent risk of dementia outcomes. Use of data linkage would represent the most cost-efficient approach,9 though as noted in Chapter ??, there is the potential for non-differential misclassification when defining dementia outcomes using EHRs, unless there was additional case validation using research criteria on top of routine diagnosis.

### 1.7.2 Evidence on vascular dementia

As highlighted by the results of the systematic review presented in Chapter ??, there is an absence of evidence on vascular dementia across the existing evidence base. While the primary studies presented here go some way towards supplementing this evidence base, future work is needed both to investigate the reasons for this evidence gap (e.g. due to an absence of primary data, challenges in the analysis of this outcome as documented in Chapter ??, or a publication bias mechanism) and to address it.

However, it seems unlikely that further analysis of observational data alone will be sufficient to investigate this outcome, as indicated by the presence of strong confounding by indication in the analysis of statins and vascular dementia (Chapter ??). Even using new methods, such as a target trial emulation approach, previous analyses have not replicated the known protective effect of statins on coronary heart disease identified by RCTs (see Section ??).10 Similarly, Mendelian randomisation studies are limited to a single one-sample analysis of HMGCR variants as a proxy for statin treatment. This study suggested a protective effect of lipid-lowering on vascular dementia, though there was a moderate risk of bias in this analysis and the effect was attenuated to the null when including a wider range of lipid-lowering variants.

Taken together, these points reinforce the need for large-scale GWAS of vascular dementia, similar to those current available for Alzheimer’s disease, to identify associated loci that future two-sample Mendelian randomisation studies could exploit. While GWAS of this outcome are likely to be methodologically challenging, given the difficult in diagnosing “pure” vascular dementia, it would be a worthwhile endeavour and allow for the assessment of genetically-driven risk factors beyond those considered in this thesis.

### 1.7.3 Preprinted evidence

In terms of evidence synthesis methodology, future work on preprinted evidence should address how to handle discrepancies between the preprinted and published versions of a paper when identifying results for inclusion in a synthesis. This will be particularly important in cases where the results are substantially different between the two versions or if a result available in the preprint manuscript is not presented in the published version.

The medrxivr tool enables programmatic searching of health-related preprints, and future work could incorporate the tool, along with software to search other literature sources, into an automated searching pipeline as part of a living systematic review.11 In addition, the ready access to preprint data afforded by the tool will enable future meta-epidemiological studies to examine factors which may influence eventual publication (e.g. significant result, geographical location/gender/career stage of first author, or some other factor). Finally, the tool will allow future research to assess the impact of peer review and editorial guidelines on manuscripts through comparison of the same paper at two stages (preprinted versus published). This approach was used to explore the impact of editorial polices on open data sharing, as discussed in Chapter ??, but could be applied to other aspects of the publication process.

### 1.7.4 Reviewing Mendelian randomisation studies

As noted in Chapter ??, methods for the inclusion of Mendelian randomisation studies in evidence synthesis exercises are not yet sufficiently developed, particularly in comparison to other study designs. Future work should aim to validate search strategies for this study design, paying particular attention to the range of terms used to define the analytical approach (e.g. genetic instrumental variable analysis). In addition, as discussed in Section ??, the lack of an established risk-of-bias tool for Mendelian randomisation studies limits subsequent analyses that requires systematic assessment of bias, such as the quantitative triangulation framework proposed here. Creation of a domain-based tool for this study design, which accounts for the differences in potential biases between one- and two-sample approaches, should be considered a priority. Finally, the inclusion of multiple two-sample Mendelian randomisation studies using identical summary statistics can falsely increase the precision of the summary effect estimate if treated as independent results. Development of best practice guidance for addressing this problem would represent a useful contribution to the field.

### 1.7.5 Systematic reviews and quantitative triangulation

Quantitative triangulation is an area ripe for future work. The approach presented here will benefit from future meta-epidemiological studies to better define the impact of bias/indirectness, as discussed in detail in Section ??. Similarly, future work should expand the framework proposed here to account for meta-biases, such as those introduced by missing evidence. Tools such as ROB-ME will enable this, but how to adjust for this analysis-level - as opposed to result-level - bias will need to be determined. Finally, the application of the framework to other causal questions will help to identify sticking points and may lead to refinement of the process.

## 1.8 Overall conclusions

To conclude, this thesis has provided new evidence concerning the role of blood lipids as a modifiable risk factor for dementia and highlighted the considerable uncertainty that still remains in relation to this causal question. In addition, it has developed new methods and tools, specifically around the inclusion of preprints in systematic reviews and the quantitative triangulation of evidence sources, which will enable future evidence synthesists to better address important causal questions.

1. Iwagami, M. *et al.* Blood cholesterol and risk of dementia in more than 18 million people over two decades: A retrospective cohort study. *The Lancet Healthy Longevity* **2**, e498–e506 (2021).

2. Gong, J., Harris, K., Peters, S. A. E. & Woodward, M. Sex differences in the association between major cardiovascular risk factors in midlife and dementia: A cohort study using data from the UK Biobank. *BMC Medicine* **19**, 110 (2021).

3. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **12**, e1001779 (2015).

4. Katikireddi, S. V., Egan, M. & Petticrew, M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. *Journal of Epidemiology and Community Health* **69**, 189–195 (2015).

5. Pérez-Gaxiola, G., Verdugo-Paiva, F., Rada, G. & Flórez, I. D. Assessment of Duplicate Evidence in Systematic Reviews of Imaging Findings of Children With COVID-19. *JAMA Network Open* **4**, e2032690 (2021).

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

7. Group, H. P. S. C. Effects on 11-year mortality and morbidity of lowering LDL cholesterol with simvastatin for about 5 years in 20 536 high-risk individuals: A randomised controlled trial. *The Lancet* **378**, 2013–2020 (2011).

8. Ford, I., Murray, H., McCowan, C. & Packard, C. J. Long-Term Safety and Efficacy of Lowering Low-Density Lipoprotein Cholesterol With Statin Therapy. *Circulation* **133**, 1073–1080 (2016).

9. Llewellyn-Bennett, R. *et al.* Post-trial follow-up methodology in large randomised controlled trials: A systematic review. *Trials* **19**, 298 (2018).

10. 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).

11. Elliott, J. H. *et al.* Living Systematic Reviews: An Emerging Opportunity to Narrow the Evidence-Practice Gap. *PLOS Medicine* **11**, e1001603 (2014).