1

Systematic review of all evidence available on the association between blood lipids (and treatment) and dementia outcomes

1.1 Lay summary

Systematic reviews are a type of research that aim to use all existing evidence to provide the best answer to an important research question. They do this by finding and combining the results from many related primary research studies. Reviews involve multiple steps including: searching of existing studies; assessment of the studies against predefined inclusion criteria; collection of data from each study; assessment of each study's methods.

This chapter presents a systematic review of primary studies that have examined the relationship between the levels of blood lipids (such as cholesterol and triglycerides), and treatments that change these levels, and dementia.

There were 127 primary studies that contained information on this relationship. I found that statins reduce the risk of Alzheimer's disease, but had no effect of vascular dementia. Lipids were not associated with any outcome. The methods used in some of the primary studies meant that I was less confident in the accuracy of their results.

The use of the results of this review in subsequent chapters is dicussed.

1.2 Introduction

In this chapter, I describe a comprehensive systematic review of the relationship between blood lipid levels, and treatments that modify them, and the subsequent risk of dementia and related outcomes. This analysis sought to address two specific aims.

Firstly, as discussed in the Introduction to this thesis (Section **??**), several diverse forms of evidence on the relationship of lipids and dementia exist. These include randomised controlled trials, observational studies of different analytical design (cohort, case-control, etc), and Mendelian randomisation studies. However, based on a scoping review of existing literature, no previous evidence synthesis exercise has attempted to examine the association of lipids/statins with dementia outcomes across these distinct evidence types. Collating these diverse evidence sources is important, as if the observed association between lipids and dementia is constant across them, it increases our confidence in the association. As such, the primary aim of this analysis was to systematically review all available literature, regardless of study design.

Secondly, I explicitly sought to include health-related preprint servers as a potential evidence source in this review, as they are infrequently considered by evidence syntheists but may contain important unpublished studies. As an adjunct analysis to the systematic review presented in this chapter, I sought to quantify the additional evidential value of including preprints. This inclusion of preprint serves makes use of the preprint search tool presented in Chapter **??**.

The results of this review are used to guide the primary analysis presented in Chapter **??**, in addition to forming a key evidence source used in the triangulation exercise presented in Chapter **??**.

1.3 Methods

1.3.1 Protocol

A pre-specified protocol for this analysis was registered using the Open Science Framework, and is available for inspection.[2] Deviations from this protocol are detailed in the relevant sections.

1.3.2 Contributions

In line with best-practice guidance, secondary reviewers were used to check the accuracy of screening, data extraction and risk-of-bias assessment processes. Due to the scale of the project, this systematic review was performed in conjunction with a team of secondary reviewer. These included Alexandra MacAleenan, Athena Sheppard, Matthew Lee and Lena Schmidt. In addition, Sarah Dawson, and information specialist, provided input to the design of the search strategy.

1.3.3 Search strategy

I systematically search electronic bibliographic databases to identify potentially relevant entries (hereafter referred to as "records"). The search strategy used in each database was developed in an iterative manner using a combination of free text and controlled vocabulary (MeSH/EMTREE)[3] terms to identify studies which have examined the relationship between blood lipids levels and dementia, incorporating input from an information specialist. The strategy included terms related to lipids, lipid modifying treatments, and dementia and its sub-types, and was designed for MEDLINE before being adapted for use in the other bibliography databases listed. An outline of the general strategy is presented in the Table 1.1 below and the full search strategies for each database are presented in Appendix **??**.

Table 1.1: Summary of systematic search by topic. The full search strategy including all terms and the number of hits per term is included in Appendix **??**.

No.	Concept
1	Dementia/Mild Cognitive Impairment
2	Lipids
3	Lipid-modifying treatments
4	1 AND 2
5	1 AND 3
6	4 OR 5
7	Animals NOT (Animals AND Humans)
8	6 NOT 7
9	Observational filter
10	Randomised controlled trial (RCT) filter
11	Mendelian randomisation/Instrumental variable filter
12	OR/ 9-11
13	8 AND 12

For all topics, search queries were comprised of relevant free text & controlled vocabulary terms.

When searching the bibliographic databases, study design filters were employed to try and reduce the screening load. To ensure that the study design filters are not excluding potentially relevant records, a random sample of 500 records identified by the main search but excluded by the filters (defined as "8 NOT 12" in Table 1.1) was screened.

The following databases were searched from inception onwards: Medline, EMBASE, Psychinfo, Cochrane Central Register of Controlled Trials (CENTRAL), and Web of Science Core Collection. As the contents of the Web of Science Core Collection can vary by institution, the specific database searched via this platform are listed in Appendix **??**. I also search clinical trial registries, for example ClinicalTrials.gov, to identify relevant randomized controlled trials.

In addition, I searched the bioRxiv and medRxiv preprint repositories using the tool developed in Chapter **??** to identify potentially relevant preprinted studies.

Grey literature was also searched via ProQuest, OpenGrey and Web of Science Conference Proceedings Citation Index, while theses were accessed using the Open Access Theses and Dissertations portal. In addition, the abstracts list of relevant conferences (e.g. the proceedings of the Alzheimer's Association International Conference, published in the journal Alzheimer's & Dementia) were searched by hand. Finally, the reference lists of included studies were searched by hand while studies citing included studies was examined using Google Scholar (forward and reverse citation searching or "snowballing").

1.3.4 Study selection

Records were imported into Endnote and de-duplicated using the method outlined in Bramer et al. (2016).[4] In summary, this method uses multiple stages to identify potential duplicates, beginning with automatic deletion of records matching on multiple fields ("Author" + "Year" + "Title" + "Journal"), followed by manual review of less similar articles (e.g. those matched based on the "Title" field alone).

Following deduplication of records, screening (both title/abstract and full-text) was performed using a combination of Endnote, a citation management tool,[5] and Rayyan, a web-based screening application.[6] Title and abstract screening to remove obviously irrelevant records was performed primarily by me, with a random sample of excluded records being screened in duplicate to ensure consistency with the inclusion criteria.

Similarly, I completed all full-text screening, with a random ~10% being screened in duplicate by a second reviewer. In addition, any records identified I identified as being difficult to assess against the inclusion criteria were screened in duplicate. Reasons for exclusion at this stage were recorded. Disagreements occurring during either stage of the screening process were resolved through discussion with a senior

colleague. A PRIMSA flow diagram was produced to document how records moved through the review.[7]

The criteria used to assess eligibility are presented in the subsequent sections.

Inclusion criteria

I sought to include studies that examine the relationship between blood lipid levels (or any specific lipid fraction, including total cholesterol, HDL, LDL, and triglycerides) and risk of incident dementia/MCI. Eligible study designs include randomized controlled trials and non-randomized observational studies of lipid modifying treatments, longitudinal studies examining the effect of increased/decreased blood lipid levels, and genetic instrumental variable (Mendelian randomization) studies examining the effect of genetically increased/decreased blood lipid levels.

Participants were free (or assumed to be free) of dementia/MCI at baseline. Studies of any duration were included to allow for exploration of the effect of length of follow-up on the effect estimate using meta-regression. No limits were placed on the sample size of included studies.

Eligible studies defined dementia according to recognised criteria, for example the International Classification of Diseases (ICD), National Institute of Neurological Disorders and Stroke Association-Internationale pour la Recherche en l'Enseignement en Neurosciences (NINDS-AIREN), or Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria.

Bit on inclusion of EHR, where it is unclear how dementia diagnosed.

For MCI, eligible studies are those that attempted state a definition for diagnoses of MCI (e.g. an adapted version of the Petersen criteria)[8] and create ordinal groups of patients (e.g. no dementia or dementia/MCI/dementia) based on this definition.

No limitations were imposed on publication status, publication date, venue or language, although sufficiently detailed reports of the studies to be able to examine their methods were required for inclusion.

Exclusion criteria

Case-control studies, cross-sectional studies, qualitative studies, case reports/series and narrative reviews were excluded. Studies which present no evidence of attempting to exclude prevalent cases from their analyses were also excluded. Studies that measure change in continuous cognitive measures (e.g. MoCA score) without attempt to map these scores to ordinal groups (e.g. no dementia/MCI/dementia) were excluded. Conference abstracts with no corresponding full-text publication were examined, and I contacted authors to obtain information on the study's status. Studies that are reported in insufficient detail (e.g. only in conference abstracts, new, letters, editorials and opinion) were excluded. Previous systematic reviews were not eligible for inclusion, but their reference lists were screened to identify any potentially relevant articles.

Studies with outcomes not directly related to the clinical syndrome of dementia (e.g., neuroimaging), studies implementing a "multi-domain intervention" where the lipid-regulating agent is included in each arms (e.g. for example, a study examining exercise + statins vs statins alone, but a study examining exercise + statins vs exercise alone would be included), and studies where there was no screening for dementia at baseline except if the sample was initially assessed in mid-life (i.e. below the age of 50) were excluded. Finally, studies using a dietary intervention, for example omega-3 fatty acid enriched diet, were excluded as it is difficult to disentangle the effect of other elements contained within the diet. Note, this is distinct from studies which delivered a simple tablet-based omega-3 intervention, which would have been eligible for inclusion.

1.3.5 Validation of screening process

Inter- and intra-rater reliability during the screening stage were assessed for a 10% sub-sample of records at the title and abstract screening stage. Intra-rater reliability involved a single reviewer applying the inclusion criteria to the same set of records while blinded to their previous decisions (i.e. assessment of conistency), while inter-rater reliability involved two reviewers independently screening the same set of records (i.e. assessment of accuracy).

Rater reliability was assessed using Gwet's agreement coefficient (AC1).[9] This measure of inter-rater reliability was chosen over other methods of assessing inter-rater reliability such as percent agreement (number of agreements divided by total number of assessments), as it accounts for chance agreement between reviewers but does not suffer from bias due to severely imbalanced marginal totals in the same way that Cohen's *kappa* value does.[9] Given the small number of included studies in this review as a proportion of the total number screened, this is an important characteristic.

Gwet's AC1 is defined as:

$$AC1 = \frac{observed\ agreement - chance\ agreement}{1 - chance\ agreement}$$

In reference to a two-by-two table with cells A, B, C and D, it is calculated using the following:

$$AC1 = \frac{\frac{A+D}{N} - e(\gamma)}{1 - e(\gamma)} \tag{1.1}$$

where $e(\gamma)$ is the chance agreement between raters, given as $2q(1-q)$, where

$$q = \frac{(A+C) + (A+B)}{2N} \tag{1.2}$$

How to interpret agreement co-efficients is widely debated, and while arbitary cut-off values may mislead,[12] they provide a useful rubric by which to assess inter-rater

agreement. Here, I used guidelines based on a stricter interpretation of the Cohen's *kappa* coefficient,[13] presented in Table 1.2.

Table 1.2: Suggested ranges to aid in interpretation of Gwet's AC1 inter-rater reliability metric

Kappa	**Interpretation**
0 – 0.20	None
0.21 – 0.39	Minimal
0.40 – 0.59	Weak
0.60 – 0.79	Moderate
0.80 – 0.90	Strong
> 0.90	Almost perfect

If this assessment demonstrated issues with the screening process, a larger proportion of records would have been dual-screened.

1.3.6 Data extraction

Data extraction was performed using a piloted data extraction form. Extracted items included: article metadata (year of publication, author list, journal), study characteristics (location, cohort name, exposure, outcomes, outcome criteria used), patient characteristics (age, sex, baseline cognition scores, baseline education scores), and results (exposure-outcome pairing, effect measure, effect estimate, error estimate, p-value). All data was extracted the data in the first instance and was checked for accuracy by a second member of the review team.

Combining across groups

Following best practice, where summary data was presented across two groups (e.g. age at baseline stratified by hypercholesterolemai status), the following approach was used to combine the groups

$$N = N_1 + N_2 \tag{1.3}$$

$$Mean = \frac{(N_1 M_1 + N_2 N_2)}{(N_1 + N_2)} \tag{1.4}$$

$$SD = \sqrt{\frac{(N_1 - 1)SD_1^2 + (N_2 - 1)SD_2^2 + \frac{N_1 N_2}{N_1 + N_2}(M_1^2 + M_2^2 - 2M_1 M_2)}{N1 + N2 - 1}} \tag{1.5}$$

This was implemented in a systematic manner, with the raw group data being extracted and a cleaning script written to combine the groups for analysis. where more than one group

1.4

$$N = N_1 + N_2 \tag{1.6}$$

Harmonisation of cholesterol measures

Harmonization of cholesterol measures across studies was performed, as different studies used different methods to quantify exposure, including comparing differing risks in the highest vs lowest quartiles of a lipid, using a binary classification of patients into a hypercholesterolaemia or not, categorizing lipid levels into high, middle, and low groups according to study-defined criteria, and simply treating the exposure as a continuous variable. Where necessary, lipid levels reported in $mmol/L$ were converted in mg/dL using the following formula:

$$mg/dL = mmol/L \times Z \tag{1.7}$$

where $Z = 38.67$ for total cholesterol, LDL-c and HDL-c, and $Z = 88.57$ for triglycerides. For widely-used categorises of lipids levels on the *mg/dL* scale, see Table **??** in Section **??**.

Study-fication

As part of the data extraction process, multiple resords resulting from the analysis of the same data were included and grouped into single units, hereafter called studies. This is likely in the advent of multiple papers reporting results on the same cohort, but say, at different time points. Study-fication builds out the most comprehensive accounts of the studies and results from as many published articles were applicable.

This was particularly relevant to preprints and published papers reporting the same study, which were not considered duplicate records. Instead, they were considered as different reports of the same study. This is due to the potential for the published version to offer some information that the preprint did not, and vice versa.

Following up with authors

Where data points required either for the analysis or risk-of-bias assessment but were not reported, the primary authors of the study were contacted. This approach was taken due to the potentially large impact of following up with authors on the results of the review,[14]

This was particularly important for the dose response meta-analysis, where the number of participants and the cut-offs per category were often not reported.

Converting between different exposure measures

The range of effect measures presented by studies (odds ratios, risk ratios, hazard ratios, etc) are not directly interchangeable in the context of systematic review. If the outcome is rare, odds and risk ratios approximate each other. However, hazard ratios provide a very different interpretation, taking into account time-to-event in each treatment group.

Several existing reviews do not distinguish between the types of effect measures and include all existing studies in a single meta-analysis to produce an overall effect. In adddition, there is some evidence of manipulation of effect estimates in previous reviews,(e.g. Chou, Sci Reports - at least one study disagrees with) but this is not accurately documented in the review text.

1.4.1 Risk of bias assessment

A key use of the review presented in this chapter is to identify different sources of evidence at risk of a diverse range of biases, and to contrast and compare findings across them (see Section **??** for an overview of triangulation and Section **??** for the results of this qualitative analysis). To enable this triangulation exercise, a detailed and structured risk of bias assessment formed an important part of this review.

There has been a recent movement within the evidence synthesis community from examining *methodological quality* to assessing *risk of bias*,[15,16] and thus directly evaluating the internal validity of a study. Internal validity is defined here as the absence of systematic error (or bias) in a study, which may influence its results.[17,18]

This move was prompted by a unclear definition of "methodological quality" which could include facets such as, and challenges in the comparison of results from different tools. As part of this, the community also moved from checklist or score based tools towards domain-based methods, in which different potential sources of bias in a study are assessed in order.

Additionally, results should be assessed at the result (defined as a a specific outcome at a specific timepoint) rather than the study level. For example, a study may report on the efficacy of an intervention at 6 months and two years follow-up. missing outcome data that is not an issues at 6 months may introduce bias at 2 year follow-up. In this case, assigning a bias judgement to the study as a whole hides this differential biases for each result.

In this review, domain-based tools were used to assess the risk of bias for each result in each included study. The study design-specific tools are introduced and discussed in more detail in the following sections. In addition, the tool also aim to capture the potential direction of bias for each result. Possible responses included: "Favours experimental", "Favours comparator", "Towards null", "Away from null", and "Unpredictable".

Randomised controlled trials

Risk of bias assessment in randomised controlled trials was performed using the domain-based risk-of-bias assessment tool appropriate to the study design. Randomized controlled trials were assessed using the RoB2 tool.[19]

The tool assess the risk of bias across five domains: Bias arising from the randomization process, Bias due to deviations from intended intervention, Bias due to missing outcome data, Bias in measurement of the outcome, Bias in selection of the reported result. Acceptable judgements include: low risk of bias, some concerns, high risk of bias. Each of the 5 domains contains a series of signalling questions or prompts, which guide the user through the tool. Once a domain-level judgement for each domain has been assigned, an overall judgement, using the same three levels of risk of bias, is assigned to the result.

Non-randomised studies of interventions/exposures

For non-randomised studies of interventions (NRSI), I used the ROBINS-I (Risk Of Bias In Non-randomised Studies - of Interventions) tool.[16] This tool assess the risk of bias across seven domains: Bias due to confounding, Bias due to selection of participants, Bias in classification of interventions, Bias due to deviations from intended interventions, Bias due to missing data, Bias in measurement of outcomes, and Bias in selection of the reported result. Similar to RoB 2, it has a number of prompting questions per domain, with acceptable judgements including "Low risk", "Moderate risk", "Serious risk" and "Critical risk". Ideally, observationals studies should be assessed in reference to an idealised randomised controlled trial. Further, the rare overall judgement of "Low" is considered equivalent to a randomised controlled trial.

For non-randomised studies of exposures (NRSE), I opted to use a preliminary version of the ROBINS-E tool.[20] The tool is still in development, but while the signalling questions for each domain are yet to be confirmed, the overarching risk-of-bias domains have been finalised. The motivation for this using this tool above other existing published tools such as the Newcastle-Ottowa scale (NOS).[21] was two-fold. In the first instance, as mentioned in the introduction to this section, using a domain-based tool has distinct advantages over better-developed checklist-type tools such as the NOS. In addition, using a domain-based tool for non-randomised studies of exposures enabled better comparison with risk-of-bias assessments performed for the other study designs.

Mendelian randomisation studies

At present, no formalised risk-of-bias assessment tool for Mendelian randomization studies is available. Assessment of the risk of bias in Mendelian randomisation studies was informed by the approach used in a previous systematic review of Mendelian randomisation,[22] as identified by a review of risk of bias assessments in

systematic reviews of MR studies (advance results from this review were obtained from contact with the authors.). A copy of this tool is available in Appendix **??**.

Risk of bias due to missing evidence

A recent shift towards the assessment of missing evidence due to selective non-reporting - as distinct from the selective reporting of a single result from multiple planned - is demonstrated via the forthcoming RoB-ME (Risk of Bias due to Missing Evidence in a synthesis) tool. The tool is in development stages, and as part of this review, I piloted the tool, and provided feedback to the developers.

This additional appraisal marks a departure from the registered protocol, as there was initially no intention to try and examine the risk of bias due to missing evidence. This is because the tool did not exist when the protocol was originally registered.

1.4.2 Analysis methods

An inital qualiatative synthesis of evidence was performed, summarising the data extracted from studies across

Where individual studies were deemed comparable, they were incoroporated into a quantitative analysis or "meta-analysis". Meta-analysis provides a summary or pooled effect estimate across studies.

Of note, studies were not combined across different study designs (i.e. RCTs were not combined in a meta-analysis with results from observational studies). The results from each individual analytical approach were summarised, but are compare and contrasted more fully in the triangulation exercise presented in Chapter **??**

Standard meta-analysis

Both a fixed-effect and random-effects meta-analysis model was employed to combine the different included studies. The fixed-effect method was implemented as:

$$\theta_i = \mu + u_i \tag{1.8}$$

$$weighted\ average = \frac{\sum Y_i(1/SE_i^2)}{\sum(1/SE_i^2)} \tag{1.9}$$

Dose-response analyses

Several of the included studies presented data on multiple categories of lipid levels, but provided an overall effect estimate based on a comparison of only two of these categories (e.g. for example, highest vs lowest quartile). While this allows for easy interpretation of the resulting effect estimate, it ignores any potential non-linear relationships between the exposure and outcome, in addition to discarding useful information contain in the interim groups.

Where possible, data for all exposure groups was extracted. Studies were excluded from this analysis if the number of categories was less than three, if the exposure cut-off points for for each category were not presented (e.g. if the study reports splitting participants into quartiles and comparing the highest vs lowest without giving the quartile bands in mmol/L).

A spline model was fitted to allow for a non-linear relationship, for example a U or J-shaped relationship, where low and high levels of an exposure can have an effect versus the "normal" reference dose. Reference doses were defined *a priori* as the cut-off of the "Normal"/"Optimal" categories for each fractions, as detailed in Table **??**. Under this approach, the reference dose was defined as 200 mg/dL for total cholesterol, 100 mg/dL for LDL-c, 40 mg/dL for HDL-c, and 150 mg/dL for triglycerides.

Due to the requirements for the dose response analysis, studies were excluded from this secondary analysis if they did not provide the require information: cut-off points. Where this was not reported in the study, I contacted the corresponding author to attempt to obtain the required information (see 1.4).

When the highest category was open ended (e.g LDL-c $\geqslant$ 200 mg/dL), I calculated category midpoint by assuming the width of the higest category was the same as the one immediately below it. Similarly, when, the lowest category was open-ended (e.g LDL-c $\leqslant$ 100 mg/dL), I set the lower boundary for this category to zero (though this is unlikely to occur natually, it was difficult to define).

Sensitivity analyses

I conducted a leave-one-out analysis in order to explore the impact of any given study on the results. In addition, where available, I performed meta-regression to assess whether the results varied by age or sex.

Visualisation of results

Evidence maps are useful way to explore the distribution of research cohorts included in a systematic review.[23] As such, the location of each individual study contributing to the evidence base was quantified and visualised on a world map.

Further, given the importance of visualising the potential biases of a result alongside the result itself, a new visualisation tool was designed to allow for "paired" forest plots (as recommended by the ROB2 publication).[19] This tool was developed as an adjunct to this thesis to aid in creating standard risk-of-bias figures,[24] and the "paired" forest plot functionality grew out of a collaboration with other reserachers to design a modular method for creating these plots.[25] A summary of this tool is

contained in Appendix **??**, and all forest plots presented in this analysis were created using this tool.

Assessment of added value of including preprints

Preprints are a valuable evidence source (see Introduction, Section **??**) but their inclusion in a systematic review.

As part of a study within this larger review, I also to explore the additional evidental value of including preprints in the meta-analysis, assessed using the fixed effect weight from a standard meta-analysis.

Additionally, I followed preprints up over time to investigate whether all identified preprints included in the review were subsequently published (in which case preprints provide a snapshot into the future, and a systematic review update would capture these reports) or alternatively, if some preprints were not published, then preprints provide a distinct evidence source.

1.5 Results

1.5.1 Initial search and validation of search filters

Twenty-three thousand, four hundred forty-seven records were identified through database searches.

Of the 500 random records screened to ensure the accuracy of the study design filters, no eligible records were identified. Many of those excluded by the filters were basic science studies, commentaries or educational articles.

1.5.2 Screening results

Following de-duplication, the titles and abstracts of 16,109 records were assessed for eligibility. 387 were deemed potentially eligible, and the full text records for these were accessed and screened. 1.1

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) flow diagram[7] presented in Figure 1.1, illustrates the movement of articles through the review.

To highlight the contribution of preprint archives to the review, the flow diagram delineates between those records captured through databases searches (presented on the right of the diagram) and those captured by the search tool described in the previous chapter (presented in grey on the left of the diagram).

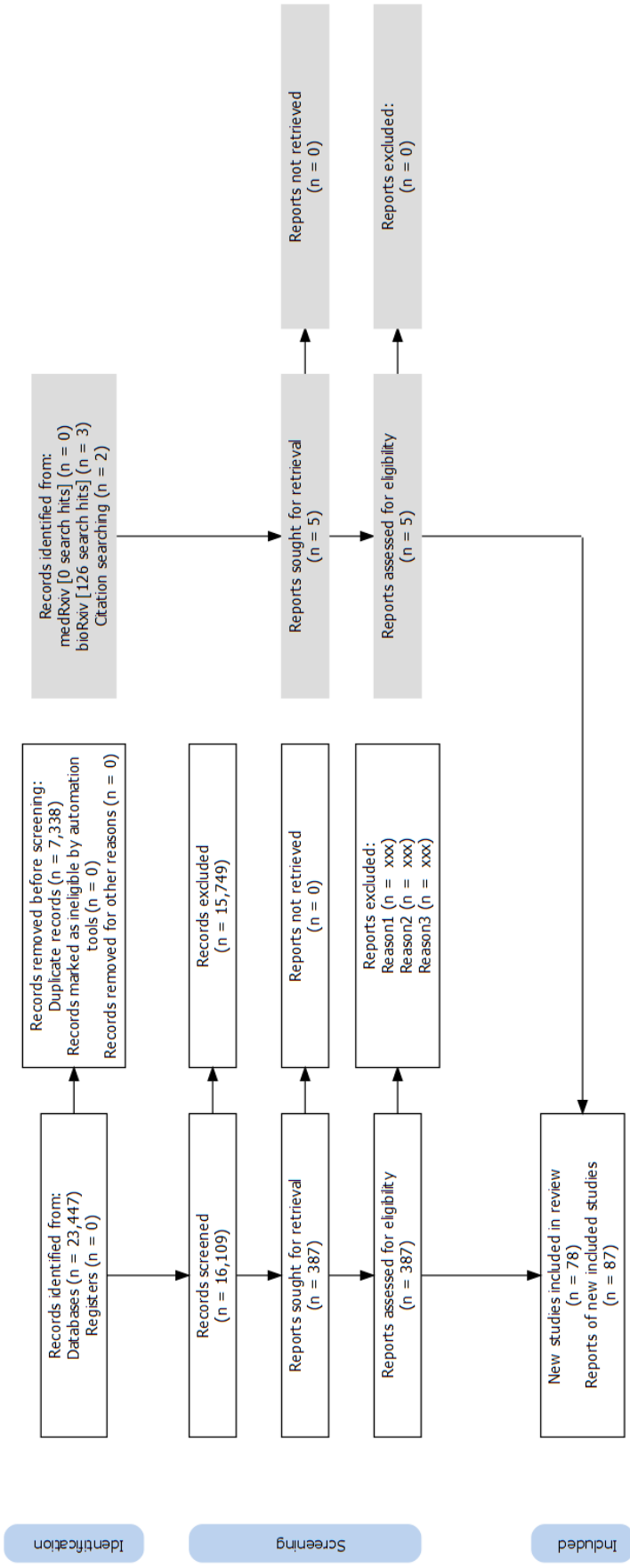

Figure 1.1: PRISMA Flow diagram illustrating how records moved through the systematic review process. The different contributions of databases and preprint servers to the review are indicated.

The values of $AC1$ were interpreted against the categories presented in Table 1.2. For the inter-rater reliability, agreement was "almost perfect" ($AC1 = 0.97$, $kappa = 0.54$, Table 1.3). Similarly for intra-rater reliability, agreement was "almost perfect" ($AC1 = 0.99$, $kappa = 0.65$, Table 1.4). The discrepancy between the $AC1$ and *kappa* coefficients illustrates the sensitivity of *kappa* to imbalanced marginals, caused in this sample by a heavy imbalance towards exclusion.[26]

Table 1.3: Inter-rater agreement on a subset of records, indicating high accuracy.

		Initial screening decision		
		Exclude	**Include**	**Total**
Second reviewer decision	**Exclude**	1244	9	1253
	Include	26	22	48
	Total	1270	31	1301

Table 1.4: Intra-rater agreement on subset of records, indicating high consistency.

		Initial screening decision		
		Exclude	**Include**	**Total**
Same reviewer decision	**Exclude**	1266	14	1280
	Include	4	17	21
	Total	1270	31	1301

Those records which were excluded in the initial screening, but were included by the second reviewer (n=26, Table @ref(tab: agreeInter-table)) were investigated. This discrepancy between the two reviewers was explained in all cases by differing interpretations of the inclusion criteria, specifically around the definition of cognitive decline versus mild cognitive impairment, and the definition of eligible lipids fractions.

1.5.3 Characteristics of included studies

Following full-text screening, **XXXX** studies met the criteria for inclusion in the review. Table 1.5 presents a summary of the characteristics of each study.

Of note, several studies provided evidence on a lipid fraction as part of a wider study on Mets. Raffatin/Ng

Table 1.5: Characteristics of included studies, stratified by study design.

Study	Location	N	Age at baseline	Female (%)	Exposures	Outcomes	Diagnostic criteria
RCTS							
HPS 2002	United Kingdom	20536	>70	24.8	Simvastatin	Dementia	NR
JUPITER 2009	Multiple	17902	66(median)60-71(IQR)	38	Rosuvastatin	Dementia	NR
NRSI							
Ancelin 2012	France	7056	NA	67	Fibrate; Statin	Dementia; AD	DSM-IV; NINCDS-ADRDA
Arvanitakis 2008	United States	929	74.9	68.7	Statin	AD	Consortium to Establish a Registry for Alzheimer's Disease (CERAD)
Bettermann 2012	United States	3069	78.6(3.3)	46.2	Statin; Non-statin LRA	Dementia; AD; Vascular component	Consensus panel - criteria not reported
Beydoun 2011	United States	1604	57.6(18.4)	38.5	Statin; TC	Dementia	DSM-III-R
Chao 2015	Taiwan	256265	73.2(7.4)	50.3	Statin	Non-vascular dementia	ICD-9
Chen 2014	Taiwan	18100	67 (8.6)	47.9	Statin	Dementia; AD; Non-AD	ICD-9
Chitnis 2015	United States	8062	74.47(9.21)	53.04	Statin	Dementia	ICD-9
Chou 2014	Taiwan	33398	>60	53.9	Statin	Dementia; AD; VaD; Non-vascular dementia	ICD-9
Chuang 2015	Taiwan	123300	54 (13)	49.1	Statin	Dementia	ICD-9
Cramer 2008	United States	1674	70 (6.8)	58	Statin	Dementia/CIND	DSM-IV
Gnjidic 2016	Sweden	2056	>60	NA	Statin	Dementia	DSM-IV
Haag 2009	Netherlands	6992	69.4(9.1)	60	Statin; Non-statin LRA	AD	NINCDS-ADRDA

Table 1.5: Characteristics of included studies, stratified by study design. *(continued)*

Study	Location	N	Age at baseline	Female (%)	Exposures	Outcomes	Diagnostic criteria
Hendrie 2015	United States	974	76.6(4.9)	69.7	Statin	Dementia; AD	DSM-IV; Consortium to Establish a Registry for Alzheimer's Disease (CERAD)
Hippisley-Cox 2010	United Kingdom	2004692	46 (14)	51	Statin	Dementia	EHR codelist
Jick 2000	United Kingdom	1364	50-89	61	Statin; Non-statin LRA	Dementia	EHR codelist
Li 2004	United States	2356	75.1(6.1)	59.8	Statin; Non-statin LRA	Dementia; AD	DSM-IV; NINCDS-ADRDA
Li 2010	United States	3392	75 (6.2)	59	Statin	AD	NINCDS-ADRDA
Liao 2013	NR	5221	NA	NA	Statin	Dementia	NR
Liu 2019	Taiwan	2012	74 (7.5)	NA	Statin	Dementia	ICD
Pan 2018	Taiwan	14807	65 (13)	43	Statin	Dementia	ICD-9
Parikh 2011	United States	377838	75.53(6.07)	2	Statin	Dementia	ICD-9
Rea 2005	United States	2798	NA	NA	Statin; Non-statin LRA	Dementia; AD; Mixed; VaD	NINCDS; NINCDS-ADRDA; Combination; State of California Alzheimer's Disease Diagnostic and Treatment Centers
Redelmeier 2019	Canada	28815	76	61.3	Statin	Dementia	ICD-9
Reitz 2004	United States	1168	78.4(6.2)	68.3	Statin; TC; Non-HDL-c; HDL-c; TG; LDL-c	VaD; AD	Cohort criteria; NINCDS-ADRDA
Smeeth 2009	United Kingdom	729529	50	40-81	Statin	Dementia; AD; Non-AD	EHR codelist
Solomon 2010	Finland	17597	68 (5.8)	57	Statin	Dementia	EHR codelist
Sparks 2008	United States	2068	75 (3.8)	54	Statin	AD	NINCDS-ADRDA

Table 1.5: Characteristics of included studies, stratified by study design. *(continued)*

Study	Location	N	Age at baseline	Female (%)	Exposures	Outcomes	Diagnostic criteria
Szwast 2007	United States	1416	77.3(5.3)	69.3	Statin	Dementia	DSM-IV
Yang 2015	Taiwan	45973	82 (5.3)	48	Statin; Fibrate; LRA (exlc. statin + fibrates)	Dementia	ICD-9
Zamrini 2004	United States	3397	73	0	Statin	AD	ICD-9
Zandi 2005	United States	3308	NA	NA	Statin; Non-statin LRA	Dementia; AD	DSM-III-R; NINCDS-ADRDA
NRSE							
Ancelin 2013	France	7053	74 (5.3)	61.1	Hypercholesterolemia	Dementia; AD	DSM-IV; NINCDS-ADRDA
Batty 2014	United Kingdom	103764	47.3(18.1)	55	Hypercholesterolemia; Non-HDL-c	Dementia	ICD
Benn 2017	NA	111194	56(median)46-66(range)	55	LDL-c; PCSK-9; HMGCR	AD; VaD; Dementia	ICD; ICD-10; NA
Beydoun 2011	United States	1604	57.6(18.4)	38.5	Statin; TC	Dementia	DSM-III-R
Bruce 2017	Australia	217	63.6(8.4)	45.6	TC; HDL-c; TG	Dementia	NA
Chiang 2007	Taiwan	785	58 (7.4)	41.4	TC; TG	Dementia; AD; VaD	ICD-9; NR
Dodge 2011	United States	822	71.6(4.7)	64.4	Hypercholesterolemia	Dementia; AD	DSM-III-R; Consortium to Establish a Registry for Alzheimer's Disease (CERAD)
Forti 2010	Italy	749	73 (6.1)	53	TG; Hypercholesterolemia	Dementia; AD; VaD	DSM-IV; NINCDS-ADRDA; NINCDS-AIREN
Gottesman 2017	United States	15407	54.2(5.8)	55	TC	Dementia	Combination
Gustafson 2012	Sweden	NR	NA	100	TC	AD	NR

Table 1.5: Characteristics of included studies, stratified by study design. *(continued)*

Study	Location	N	Age at baseline	Female (%)	Exposures	Outcomes	Diagnostic criteria
Hayden 2006	United States	3308	74.0(6.4)	58.2	Hypercholesterolemia	Dementia; AD; VaD	DSM-III-R; NINCDS-ADRDA; NINCDS-AIREN
Kimm 2011	South Korea	848505	53 (9.3)	42.2	TC	AD; VaD; Dementia	ICD-10
Kivipelto 2001	Finland	1499	50.4(6.0)	62	Hypercholesterolemia	AD	NINCDS-ADRDA
Kivipelto 2005	Finland	1449	50.6(6.0)	62	Hypercholesterolemia	Dementia	DSM-IV
Kuo 2015	Taiwan	67066	62.1(11.4)	48.4	Hypercholesterolemia	Dementia	ICD-9
Li 2005	United States	2141	74.9(5.9)	60.5	TC; HDL-c	Dementia; AD	DSM-IV; NINCDS-ADRDA
Mainous 2005	United States	6558	NA	NA	Hypercholesterolemia	Dementia; AD	ICD-9
Mielke 2005	Sweden	382	NA	70	TC; TG	Dementia	DSM-III-R
Mielke 2010	France	1460	38-60	100	Hypercholesterolemia; TC	Dementia; AD	DSM-III-R; NINCDS-ADRDA
Mielke 2012	United States	99	74 (2.5)	100	TC; HDL-c; TG	Dementia; AD	DSM-IV; NINCDS-ADRDA
Muller 2007	United States	542	NA	NA	TG; HDL-c	AD; Dementia	NINCDS-ADRDA; DSM-IV
Noale 2013	Italy	5632	71.3(5.3)	56.3	TG; Hypercholesterolemia	Dementia	DSM-III-R
Notkola 1998	Finland	444	40-59	0	Hypercholesterolemia	AD	Combination
Peters 2009	Mutliple	3336	>80	60.4	TC; HDL-c	Dementia	DSM-IV
Raffaitin 2009	France	7087	73.4(4.9)	61	TG; Hypercholesterolemia	Dementia; AD; VaD	DSM-IV; NINCDS-ADRDA; Combination
Rantanen 2017	Finland	3309	42(39–46)	0	TC; Hypercholesterolemia	Dementia; AD; VaD	NA
Reitz 2004	United States	1168	78.4(6.2)	68.3	Statin; TC; Non-HDL-c; HDL-c; TG; LDL-c	VaD; AD	Cohort criteria; NINCDS-ADRDA

Table 1.5: Characteristics of included studies, stratified by study design. *(continued)*

Study	Location	N	Age at baseline	Female (%)	Exposures	Outcomes	Diagnostic criteria
Reitz 2010	United States	1130	75.7(6.3)	65.7	TC; HDL-c; LDL-c	AD	NINCDS-ADRDA
Ronnemaa 2011	United States	2268	49.6(0.6)	0	Hypercholesterolemia	AD; VaD; Dementia	NINCDS-ADRDA; ADDTC; DSM-IV
Schilling 2017	France	9294	73.8(5.3)	61	TG; HDL-c; LDL-c; TC	Dementia; AD; Mixed	DSM-IV; NINCDS-ADRDA; NINCDS-AIREN
Solomon 2007	Finland	1449	50.4(6.0)	621	Hypercholesterolemia	Dementia	NA
Solomon 2009	United States	9844	43 (1.7)	54	TC	AD; VaD	ICD-9
Strand 2013	Norway	48793	42.6(4.3)	49	TC	Dementia; AD	ICD
Svensson 2019	Japan	781	54.1(5.6)	NA	HDL-c; Hypercholesterolemia	Dementia	DSM-IV
Tan 2003	United States	1026	76.1(5.3)	63	TC; HDL-C	AD	NINCDS-ADRDA
Tynkkynen 2016	Finland	13725	48.4(13.3)	51.6	HDL-c	Dementia; AD	ICD-10
Tynkkynen 2018	Multiple	22623	57 (9.2)	47	TC; HDL-c; LDL-c; TG	Dementia; AD	ICD-10
Wang 2012	Taiwan	1230400	60 (13)	52	Hypercholesterolemia	AD	ICD-9
Whitmer 2005	United States	8845	68 (2.6)	53.7	Hypercholesterolemia	Dementia	ICD-9
Yoshitake 1995	Japan	828	74 (5.9)	59.5	TC; TG; HDL-c; LDL-c	VaD; AD	NINCDS-AIREN; NINCDS-ADRDA
Zimetbaum 1992	United States	350	79(75-85)	64.5	TC; HDL-c; LDL-c; TG	Dementia	DSM-III-R
MR							
Andrews 2019	NA	NA	NA	NA	HMGCR; TC; LDL-c; HDL-c; TG	AD	NA
Benn 2017	NA	111194; 54162	NA	NA	LDL-c; PCSK-9; HMGCR	AD; VaD; Dementia	ICD; ICD-10; NA

Table 1.5: Characteristics of included studies, stratified by study design. *(continued)*

Study	Location	N	Age at baseline	Female (%)	Exposures	Outcomes	Diagnostic criteria
Burgess 2017	NA	21165	NA	NA	LDL-c; HDL-c; TG	AD	NA
Mukherjee 2013	NA	54162	76 (7.9)	59	HDL-c; LDL-c; TG	AD	NA
Ostergaard 2017	NA	54162	NA	NA	TC; HDL-c; LDL-c; TG	AD	NA
So 2017	NA	NA	NA	NA	NA	NA	NA
Zhu 2018	NA	54162	NA	NA	HDL-c; LDL-c; TG	AD	NA

The majority of studies were non-randomised, with only one included randomised controlled trials (the Heart Protection Study) and a relative small number of Mendelian randomisation studies. Of the non-randomised studies, many were assessments of the relationship with statis use.

Of the studies examining lipid levels, by far the most common fraction assessed was total cholesterol. The association with triglycerides levels was not consider in many papers.

The vast majority of studies examined either all-cause dementia or Alzheimer's disease, with only a small proportion examining vascular dementia.

Many studies using electronic health records as their data source did not accurately report the diagnostic criteria used to identify cases, and of those that did, none sought to validate the accuracy of their defined code lists.

Many of the studies were conducted in electronic health record databases, which have a number of advantages over smaller more traditional cohort studies.

As shown in Figure 1.2, the majority of study cohorts were based in the developed world.

Figure 1.2: Geographical distribution of study cohorts

1.5.4 Risk of bias

Many of the included results were assessed to be at high risk of bias.

The one randomised controlled trial was deemed to be a low risk of bias,

Several synthesis were also assessed to be at risk of bias due to missing evidence.

A key issue identified relating to bias due to missing data was the potential for the search to miss relevant Mendelian randomisation studies which had examined the association between Alzhiemer's disease and multiple risk factors. This is discussed further in Section 1.6.4.

Haag is a good example of potentially missing results - number of cases of other outcoems reported, but no analysis performed.

1.5.5 Statins

In the two included RCTs, statin use was not associated with dementia outcomes

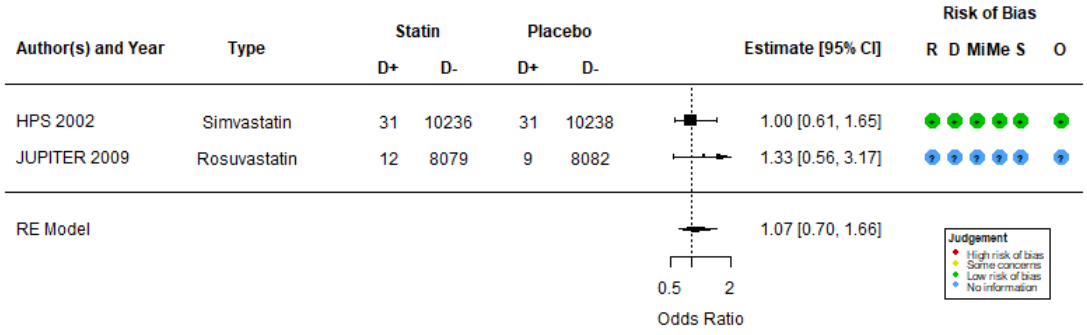

Figure 1.3: Random effects meta-analysis of statins on all-cause dementia:

In NRSI, statin use was associated with a reduce risk of all-cause dementia (OR: 0.76, 95%CI: 0.66-0.88, Figure 1.4).

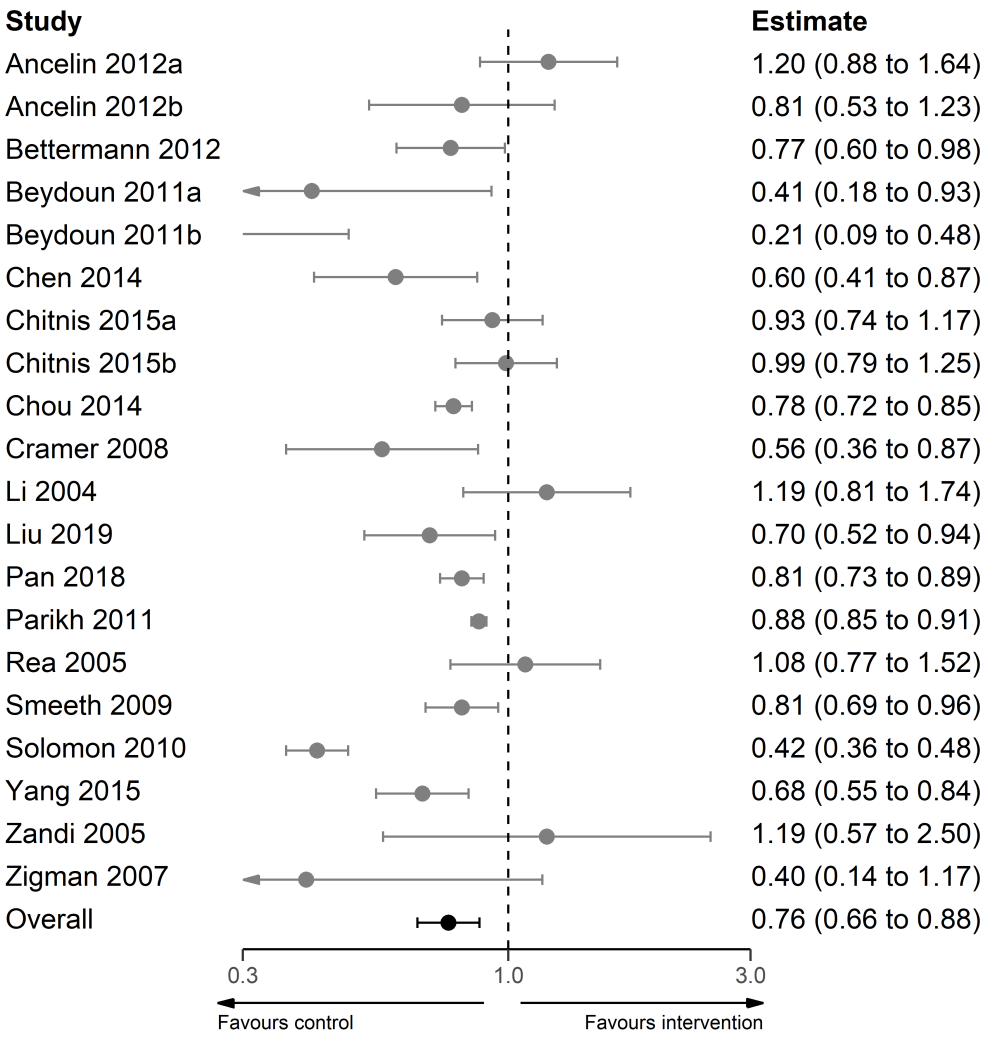

Figure 1.4: Forest plot showing effect of statins:

In Mendelian randomisation studies, a similar blah blah blah

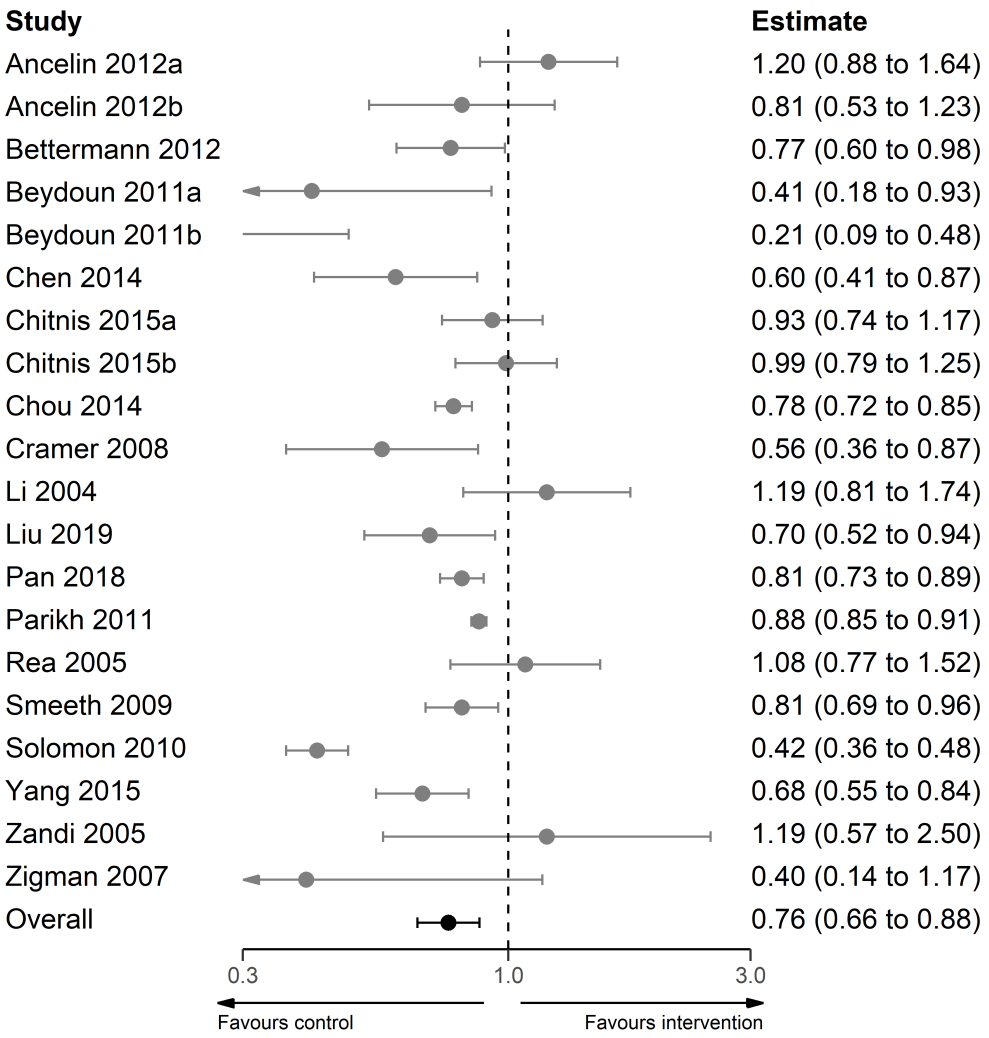

Figure 1.5: Forest plot showing effect of statins:

1.5.6 Hypercholesterolemia and lipid levels

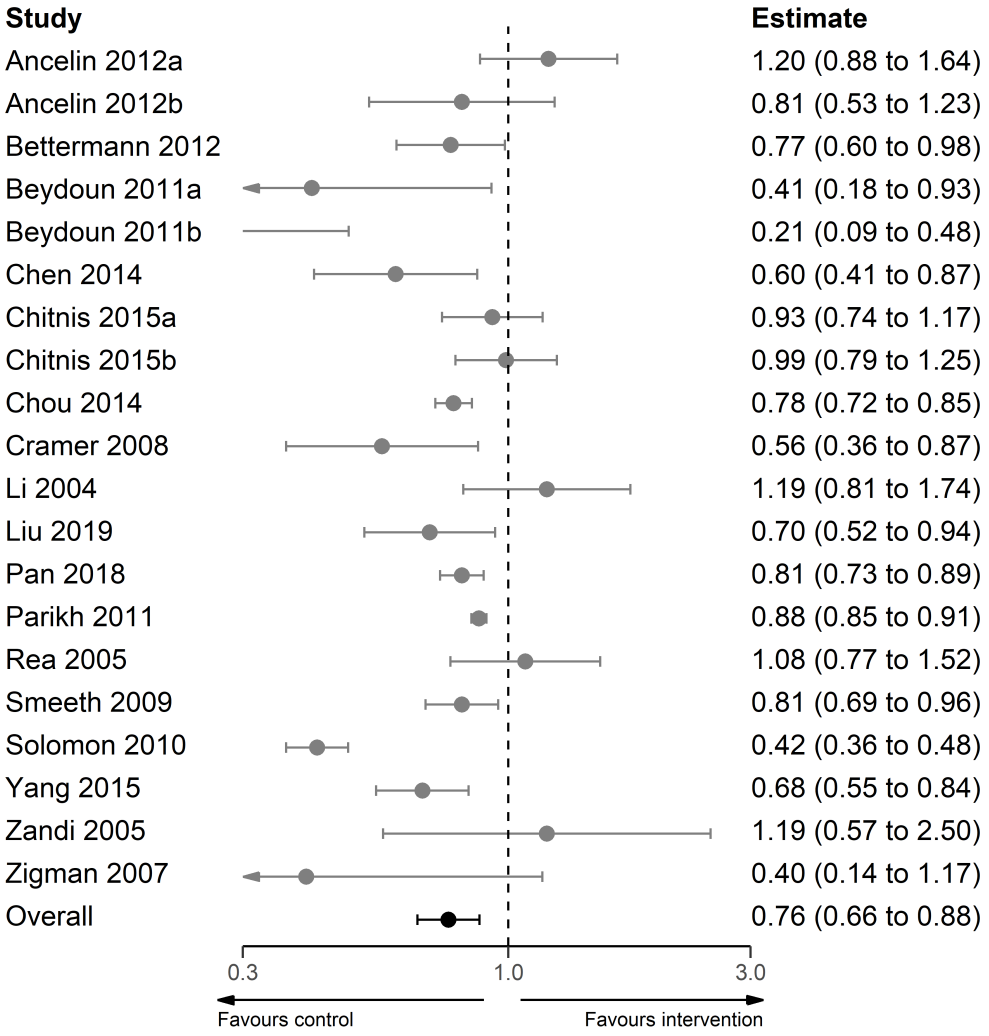

Figure 1.6: Forest plot showing effect of statins:

Several studies presented results per SD change in a lipid fraction.

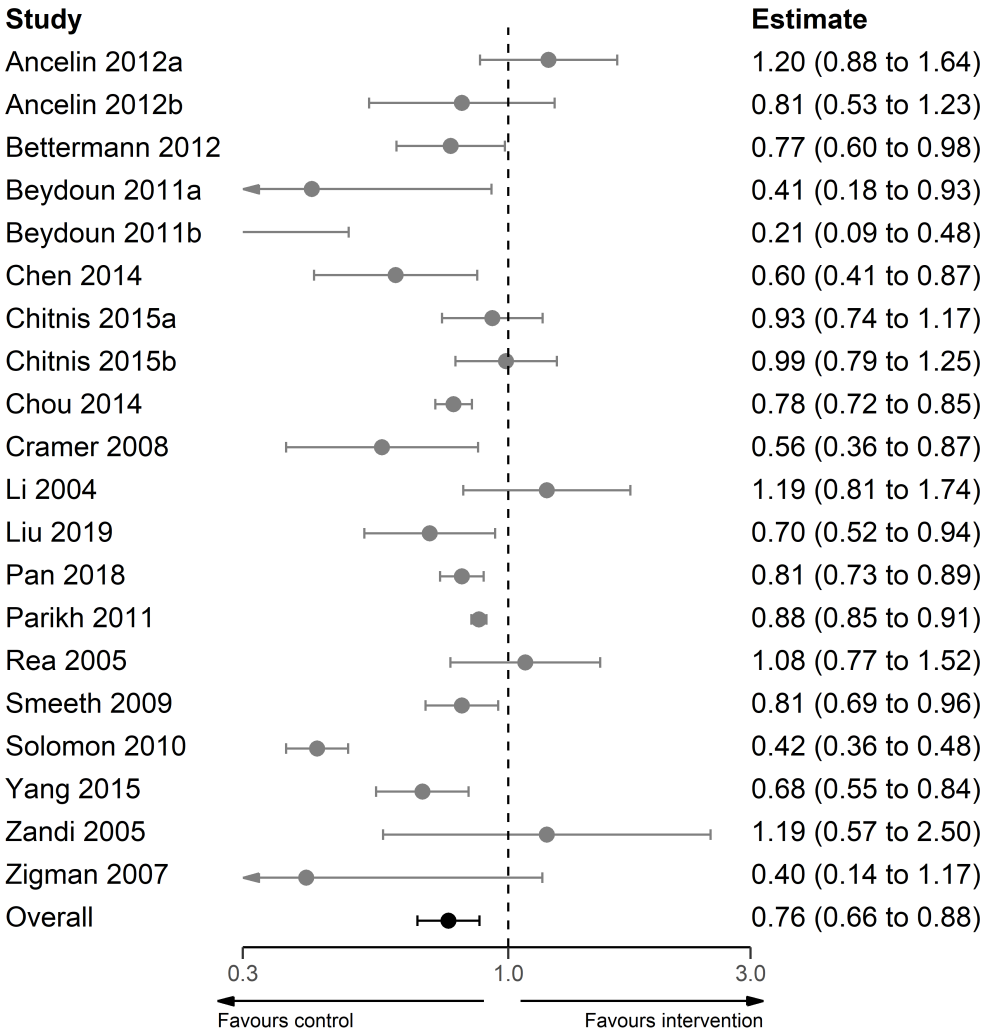

Figure 1.7: Forest plot showing effect of statins:

1.5.7 Dose response meta-analysis of lipid levels

Several studies were excluded from the dose-response meta-analysis, as the number of cases/controls per dose group could not be calculated and the corresponding author for the study did not respond to clarification requests. The results from the dose response analysis can be seen in Figure 1.8.

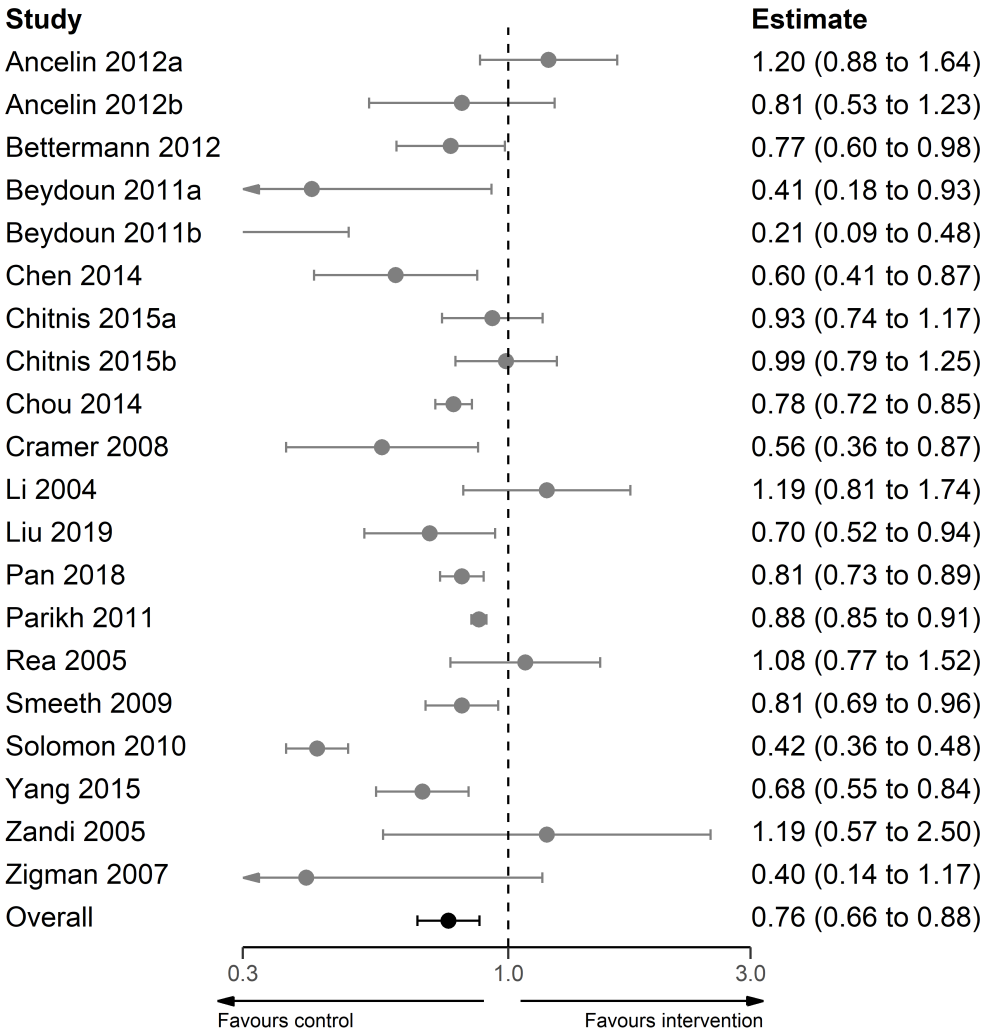

Figure 1.8: Forest plot showing effect of statins:

Planned to investigate effect of sex and age at baseline.

Plus other such as education level and baseline cognitive scores, but data were either not reported for most studies or when reported, were too diverse to synthesize.

1.5.8 Publication bias

There was little evidence of publication bias across the evidence base (Figure 1.9).

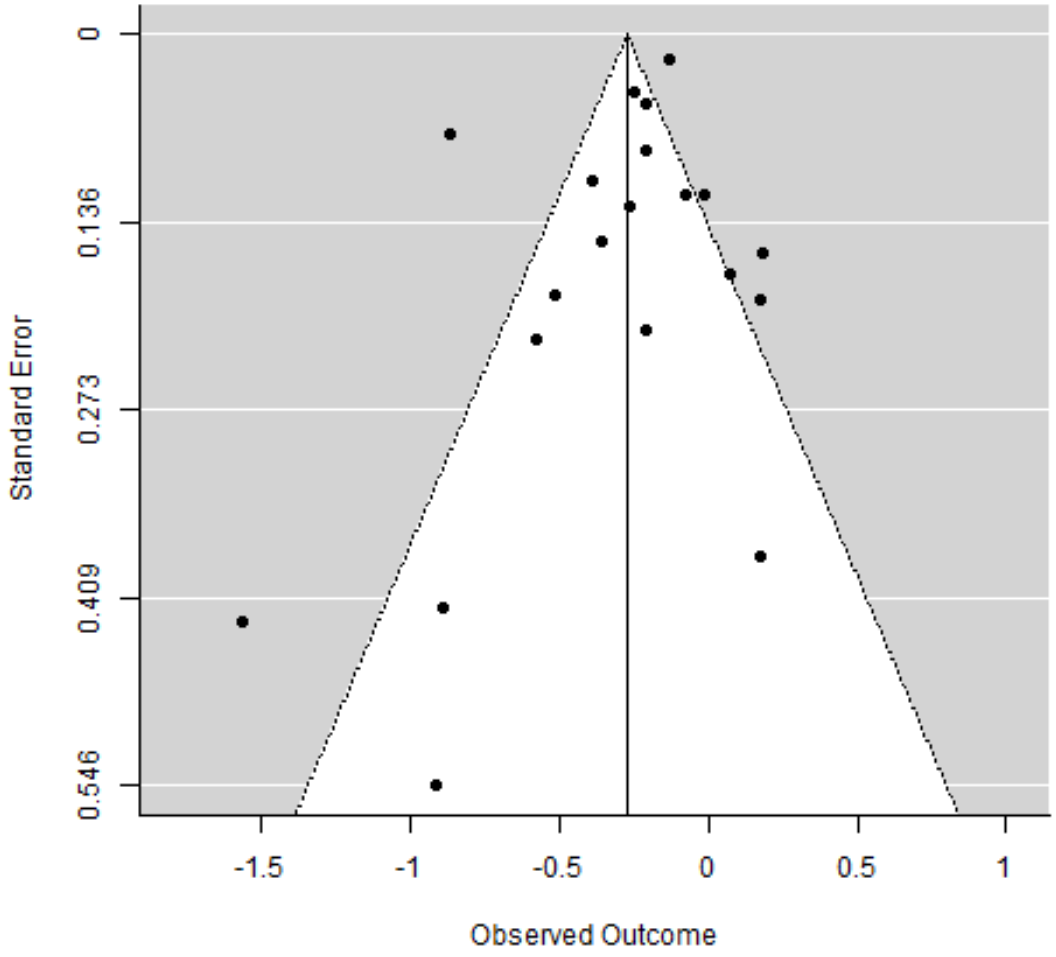

Figure 1.9: Funnel plot of results examining the relationship between statins and any dementia

1.5.9 Added evidental value of including preprints

I had intended to assess the added evidental value of including preprints as a literautre source.

As show in Figure **??**, the number of results returned by the preprint searching

was not substantial (bioRxiv = 256, medRxiv = 0).

No unique preprints were identified as part of this review, with the published version of the few/sole relevant preprint(s) being captured by the main search.

1.6 Discussion

1.6.1 Summary of findings

There was some evidence of protective effect of statins on all-cause and Alzheimer's disease dementia when looking at solely at observational studies. This finding was not supported by evidence from the single available RCT, or by studies that emulated statin treatment using a genetic proxy, suggesting that these findings are a

In addition, there were some concerns over the potential for estimates to be missing from the meta-analysis of observational studies not at random, given the preferential reporting of

Some evidence that age has impact on observed lipid-dementia relationship - use to link

This review has presented a summary of the available evidence on the association between lipids, and treatments that modify lipids such as statins, and the subsequent risk of dementia.

The distribution of evidence between analytical designs is to be expected. Randomised controlled trials of dementia are particularly challenging, as the long follow-up, necessary due to the long latent period of the condition, makes trials logistically challenging and financial expensive. Similarly, Mendelian randomisation is a relatively new study design, and so only appears in the literature in recent years, driven by the growth of dementia GWASs (Figure 1.10).

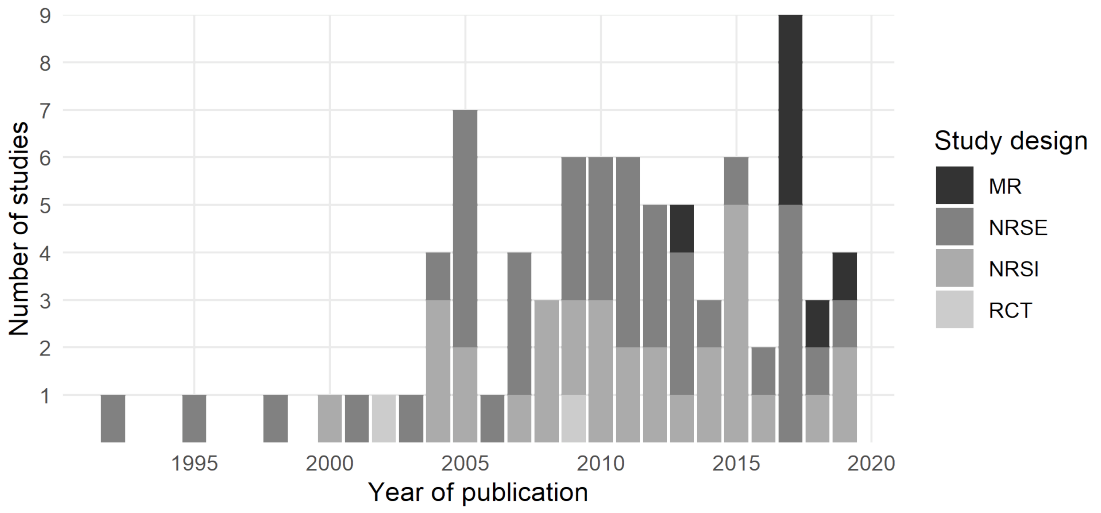

Figure 1.10: Study designs by year of publication -

A common theme across the evidence base was a lack of data on the association of vascular dementia. This is particularly interesting given that lipids and statins are primarily related to vascular disease. There is the potential that studies encountered similar difficulties in address the unexpected results observed in the CPRD analysis in Chapter **??**, likely due to confounding by indication, and so may suffer from the "file-drawer effect".[27] For vascular dementia, few Mendelian randomisation studies examined this outcome, primarily because of the absence (until recently) of genome wide association studies, which are used in two sample summary Mendelian randomisation studies.

One item of particular interest is the attenuation of any effects observed by Mendelian randomisation studies following the adjustment for/exclusion of genetic variation in the Apoe4 gene region. As covered in the discussion, ApoE4 genotype is the major risk factor for Alzheimer's disease, but is also involved in cholesterol transport.

This review did not include the commonly cited PROSPER study, which examined the effect of pravastatin on CVD risk and is regularly held up as providing no evidence for an effect of the statin on dementia outcomes. While widely cited and included in the Cochrane review on this topic,[28] only reported the mean change in a range of cognitive measures (MMSE, Stroop test, Picture-Word Learning test,

etc.) over follow-up rather than incident. While this is a useful indicator of general cognitive decline, it is not equivalent to a dementia diagnosis using a , as cognitive tests should feed into a broader diagnostic pathway (see section **??**). As such, this trial did not met the inclusion criteria for this review.

Questions over missing results - evidence from one of the conference abstract analysis pairs that a non-significant results are being suppressed.[29,30] Tie this in with

1.6.2 Previous reviews

This section will be completed once Georgia has completed her analysis, and can also be cross-references with the meta-meta analysis published is Brain Sciences recently.

While conducting this review, I identified several previous systematic reviews of this topic.[31–35] However, this review is the first to use established domain based assessments tools (for example, the RoB 2 tool for randomized controlled trials)[19] to assess the risk of bias in included studies, and explore the heterogeneity of results across different levels of risk of bias levels. Some previous reviews did assess risk of bias, but used non-domain based assessment tools, such as the Newcastle-Ottowa scale.[34,36]

However, despite these differences in methodology, the duplication of work across reviews (including this review) is substantial. In retrospect, an alternative approach to conducting a further systematic review from scratch could have been employed. Known as an umbrella review, or review-of-reviews, these studies use other systematic reviews rather than primary studies as the unit of analysis. This approach would have enable more efficient identification of relevant primary studies, to which the methods which sets this review apart from other published reviews could have been applied.

1.6.3 Inclusion of preprints

As highlighted in Section **??**, this review explicitly sought to synthesize evidence across different publication statuses (preprinted vs. published). Using the tool described in Chapter **??**, two preprint serves related to health and biomedical sciences were search as part of this review (see Appendix **??** for the code used to search the repositories). There were several relevant preprints captured by the search. The added evidential value of including these preprints was highlighted in Section **??**sys-rev-including-preprints).

Three relevant preprints were identified

medRxiv grew out of the Epidemiology and Clinical Trials categories of the bioRxiv preprint server. The small number of studies return by the searches (or the absence of any hits in the medRxiv database - see Figure 1.1) is due to the timing of the preprint searches. The searches for this review were performed in mid-July 2019, but the first medRxiv preprint was registered on 25th June 2019. As such, at the point it was searched, the medRxiv database contained only a small number of records (n=148).

Of the three identified preprints, two were subsequently published as of September 2020. This fits well with the analysis presented in Section **??** that, allowing for a two-year lag, approximately two-thirds of preprints are published.

While none of the preprints contributed uniquely to the review, this is again. Consider the example of one of the Mendelian randomisation analyses of the effect of LDl-c on Alzheimer's disease, which found no impact of LDL-c on AD following removal of APOE4. The manuscript was initially published in bioRxiv in July 2017[37] which was subsequently published in Nature Communications in January 2018, following peer-review.[38] While this study was captured by both the preprint and published searches in this review, had the searches been run within this window, the preprint would have contributed unique data to the review. This illustrates that

while it may not have aided this review, if the aim is to find the current state of the art in the topic area at the time of searching, inclusion of preprints is a necessity.

Of note, since the start of my review, inclusion of preprints in systematic reviews has now become widespread due to the role of preprint servers, in particular medRxiv, as a key evidence dissemination venue during the early stages of the COVID-19 pandemic. However, how well this adoption of preprints will transfer to other topics, where the speed of research does not put the same focus on preprinted articles, is currently unknown.

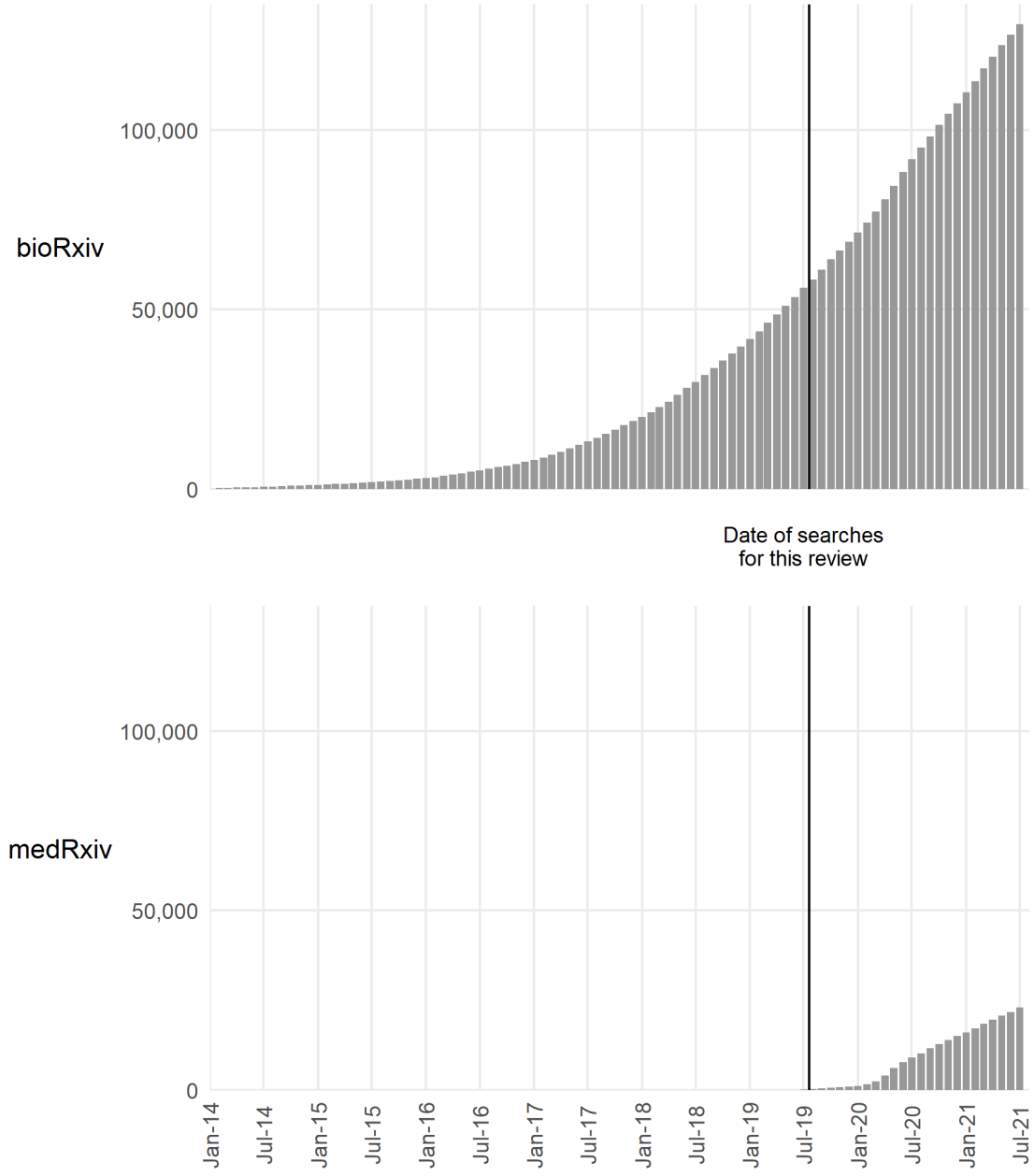

Figure 1.11: Growth of preprint repositories over time - Given the relative sizes of the preprint repositories at the time the searches for this review were conducted (bioRxiv n= 56,007, medRxiv n = 148), the relative number of hits returned by each is expected.

1.6.4 Reviewing Mendelian randomisations studies

One of the strengths of this review is it's inclusion of MR studies as a source of evidence.

Mendelian randomisation is a powerful analytical technique, using natural variation in participants genomes to (assuming the assumptions of the method are valid), though it's inclusion as an acceptable study design in this review was complicated by a number of factors.

Firstly, this study design is relatively new, particularly when compared to randomised trials or cohort studies. Figure 1.10 demonstrates that MR studies only begin to appear in the evidence base much later than NRSE/NRSI. As such, the process and tools for systematically assessing them are not as well developed, likely due to the limited availabilt of large scale GWAS datasets needed for two sample MR. A key example of this is in the absence of validated search filters for Mendelian randomisations studies. This limitation is further complicated by the varying terminology used to describe the method, particularly in the early years of it's application.

Additionally, there is currently no widely used risk-of-bias assessment tool for Mendelian randomisation studies. A recent commentary provided a checklist interpreting Mendelian randomisation studies, this guide includes reporting items in their quality checklist. While reporting quality is important, it is a separate consideration to internal validity, as discussed in Section @ref(). Similarly, a previous review of Mendelian randomisation studies used the Q-Genie tool, which was validated to assess the quality of genetic association studies in meta-analysis.[39] While this tool addresses the studies used, it does not access the additional methodological considerations of the analysis of the Mendelian randomisation analysis itself. For this review, I utilised the best available author-devised tool, sourced on a recent review of systematic reviews of Mendelian randomisation studies.

As a further stumbling block, Mendelian randomisation, particularly when using a two-sample summary data design, is a form of analysis that lends itself to multiple exposure-outcome comparisons. This is particularly relevant to the consideration of bias due to missing evidence. As an example, through snowballing and other measures, I identified at least one relevant Mendelian randomisation study that had not been identified by the search strategy.[40] On review of this paper, the search would not have been expected to find it given the absence of any lipid-related keywords in the title and abstract. The study examined the association between lipid fractions with Alzheimer's disease as one of many risk factors for the condition. Studies such as this can introduce bias into a systematic review, as it is commonly only those risk factors that show a statistically significant result that are reported in the abstract and so are captured by the search. This may bias systematic reviews, including this one, as the analysis of multiple risk factors against a single outcome within a single publication becomes more common. These studies are described as "unknown unknown's" in the context of the RoB-ME tool, and are particularly challenging (as opposed to an analysis that was insufficiently reported to be included in the statistical analysis, or the "known unknown's").

Useful future work to improve the methodology for inclusion of Mendelian randomisation studies in systematic reviews should involve the development of a validated search filter for this study design.[41,42] Alternatively, in better-resourced reviews, a dedicated search for "risk factors" and "dementia" and "Mendelian randomisation", followed by manual review of studies that look across multiple risk factors, would be advisable. This was not feasible in the context of this review, given the large number of records to be screened, even when using study design filters (n=16,109). Additionally, the value of methods that supporting the traditional bibliographic database search, such as snowballing (forwards and backwards citation chasing) and communication with relevant topic experts should not be underestimated. Finally, development of a risk-of-bias assessment tool by a panel of methodologists and analysts would be of substantial benefit.

Talk about problem with studies sharing underlying datasets in two sample MR frameworks - c.f. EHR databases, which contain the same underlying sample but use different sub-samples on account of the distinct codes/conditions and timepoints used to defined the study cohorts, MR analysis can be multiple studies using the exact same summary statistics from the same cohorts

Introduce assumptions underlying MR (with image). Higlight that most if not all included analysis initially describing a protective effect of LDL-c, which attenuates to the NULL once ApoE4 gene regions are removed. Note the region must be quite wide - talk about controversy surrounding Benn paper

Additionally many of the studies suggested a link between, but then in the sensitivity analyses or in the Discussion disclaim that when adjusted for ApoE4, any association was attenuated to the null. This point is important given one of the core assumptions of

MR are limited by the need for SNP to be present in both datasets (for TSMR).

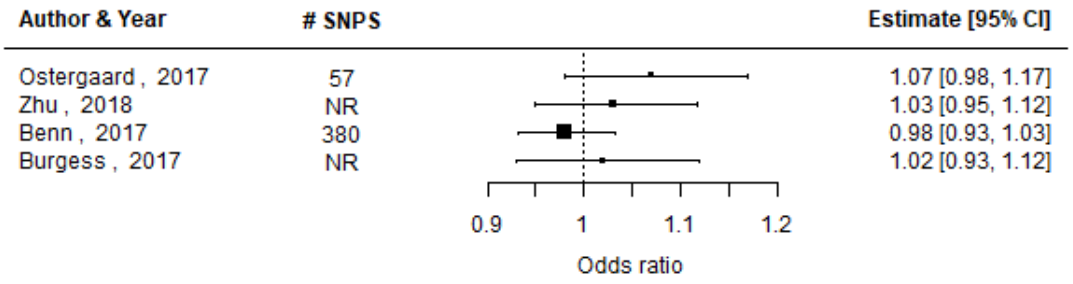

Figure 1.12: Summary of duplication of MR studies:

Recommend future large scale GWAS of other dementia outcomes, notably vascular dementia, or the use of exisitng GWASes to

A key example of the importance of . In almost all cases, MR studies examine. A clear exampe of this is Ben et al, 2017, where the ApoE variants were not sufficient identified and excluded, and the published paper detailed evidence for

a protective effect of LDL-c was identified (`estimate(0.83, .75,.92,"RR")`).[43] Following several rapid responses, the data was re-analysed excluding a larger area around ApoE4 which attenuated this finding towards the null.[44]

1.6.5 Comments on the process

As part of the reflective element of this thesis, I collated my experiences on performing a systematic review.

Protocol registration

While the protocol was registered on the Open Science Framework largely to allow for the sharing of associated documents such as the proposed search strategy, anecdotal evidence from colleagues and collaborators suggested that the findability of the protocol was limited. In hindsight, I should also have cross-posted the protocol on PROSPERO, the international prospective register of systematic reviews,[45] (which does not allow for the uploading of related files).

People management

This review provide an excellent opportunity to gain experience in managing a team of researchers. However, due to the need for dual screening and data extraction over a long period of time, a number of external researchers became involved in this review. I found the people-management aspect particularly challenging, and in retrospect could definitely have improved the process through better communication of expectations and deadlines.

Open data sharing

As discussed in Section 1.5.7, many primary studies did not report important elements, and so these could not be extracted. This limiation was compounded by the expected low response rate to requests for further information from primary authors (although, in hindsight, the form of contact used (email) has been shown to be less successful in eliciting responses from authors when compared with telephoning[46]).

While contacting authors is worthwhile, as it can substantially change the conclusion of a systematic review[47] and is not too costly to systematic reviewers,[48] a far preferable option is that the authors of primary studies readily deposit all relevant study data at the point of publication.

Based on my experience of extracting data for this review, I co-wrote an guidance article to aid primary prevention scientists in preparing and sharing their data so that it can easily be incorporated into a evidence synthesis exercise, using a trial of mindfulness interventions as an case study.[49]

Similarly, a substantial amount of time and effort has gone into making the data obtained by this review openly available to other researchers.

1.6.6 Strengths and limitations

Strengths

I believe there are four aspects where this review is distinct from those reviews already available in the published literature (as identified by):

- *Comprehensiveness:* While several reviews of this research topic exist,[31–34] 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. In addition

- *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 **??**, 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.

- *Contribution to methods work:* A large part of this review was the associated work on improving research synthesis methods. This work is detailed as relevant throughout the Chapter, often referring to additional work detailed in the . In addition this review was used to pilot an upcoming risk of bias tool

Limitations

The primary limitation of this review is that several included studies used data from EHR databases, which come with serious concerns regarding validity[50] [51,52] Relatedly, several studies which made use of electronic health record database did not report the specific code lists used, potentially introducing substantial heterogeneity between effect estimates. An empirical example of the effect of differing EHR code list is presented as part of the analysis in Chapter 4 (see Section **??**).

In addition, the fact that only a sample of records were dual screened at the title/abstract and full-text stages is a potential limitation, as there is a chance that some eligible records could have been excluded. However, evidence from assessments of inter- and intra-rater reliability indicate that is is not a major concern.

One particular limitiation with regards to the risk of bias assessment is the fact that the ROBINS-E assessments were performed without the tool being finalised. This meant that there were no signalling questions to guide the domain-level risk of bias assessment, which may have influenced the accuracy with which domain-level judgements were assigned. However, there is no published empirical evidence supporting the need for signalling questions, and assessment of inter-rater reliability across the different tools did not indicate a specific problem with the ROBINS-E assessments. In fact, low agreement was common across the tools, though this is expected based on the available literature.[53]

One further limitation is the fact that the risk of bias due to missing evidence assessment, combined with some empirical evidence that some studies were missed by the search but contained relevant studies is a definite limitation of this review (see Section @Ref(rev-discussion-MR) above for a fuller discussion of this issue with respect to Mendelian randomisations studies). Unfortunately, this is probably a common limitation across all reviews, based on the way in which increased sensitivity must be balanced with a reasonable workload.

1.7 Conclusions

In this chapter I have presented a comprehensive systematic review of the different sources of evidence available which examined the relationship between lipid levels and dementia use.

This work built on the tool introduced in the preceeding chapter (Chapter **??**sys-rev-tools-heading)), and findings from this review are used though out the subsequent

chapters: in Chapter **??**, summary of the evidence guided the choice of analysis approach, ensuring that the new analysis was at risk of a different source of bias; while in Chapter **??**, prospective cohorts identified by the review were contacted in an attempt to obtain individual participant data; finally, the cumulative effect measures calculated here are used as a key source of evidence for the triangulation exercise presented in Chapter **??**discussion-heading).

References

00 1. Cochrane, A. L. 1931-1971: A critical review, with particular reference to the medical profession. *Medicines for the year 2000* (1979).

2. McGuinness, L., Higgins, J., Ben-Shlomo, Yoav, Coulthard, L. & Smith, G. The relationship between blood lipid levels and dementia: Protocol for a systematic review and triangulation of all available evidence. *OSF* (2020). doi:10.17605/OSF.IO/VTW5Y

3. Lefebvre, C. *et al.* Section 4.4.4 -Searching for and selecting studies: Controlled vocabulary and text words. in *Cochrane Handbook for Systematic Reviews of Interventions* (eds. Higgins, J. et al.) (Cochrane, 2021).

4. Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L. & Bekhuis, T. De-duplication of database search results for systematic reviews in EndNote. *Journal of the Medical Library Association : JMLA* **104**, 240–243 (2016).

5. Hupe, M. EndNote X9. *Journal of Electronic Resources in Medical Libraries* **16**, 117–119 (2019).

6. Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews* **5**, 210 (2016).

7. Page, M. J. *et al.* The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **372**, n71 (2021).

8. Petersen, R. C. *et al.* Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology* **56**, 303–308 (1999).

9. Gwet, K. L. Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology* **61**, 29–48 (2008).

10. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* **20**, 37–46 (1960).

11. Wongpakaran, N., Wongpakaran, T., Wedding, D. & Gwet, K. L. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology* **13**, 61 (2013).

12. Brennan, P. & Silman, A. Statistical methods for assessing observer variability in clinical measures. *BMJ : British Medical Journal* **304**, 1491–1494 (1992).

13. McHugh, M. L. Interrater reliability: The kappa statistic. *Biochemia Medica* **22**, 276–282 (2012).

14. Reynders, R. M., Ladu, L. & Girolamo, N. D. Contacting of authors modified crucial outcomes of systematic reviews but was poorly reported, not systematic, and produced conflicting results. *Journal of Clinical Epidemiology* **115**, 64–76 (2019).

15. McGuinness, L. A., Higgins, J. P. T. & Sterne, J. A. C. Assessing the Credibility of Findings From Nonrandomized Studies of Interventions. *JAMA Cardiology* **3**, 905–906 (2018).

16. Sterne, J. A. *et al.* ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* **355**, i4919 (2016).

17. Campbell, D. T. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* **54**, 297–312 (1957).

18. Jüni, P., Altman, D. G. & Egger, M. Assessing the quality of controlled clinical trials. *BMJ* **323**, 42–46 (2001).

19. Sterne, J. A. C. *et al.* RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ* **366**, (2019).

20. Morgan R *et al.* A tool to assess Risk Of Bias In Non-randomised Studies - of Exposures (ROBINS-E). in *Advances in Evidence Synthesis: Special issue, Cochrane Database of Systematic Reviews* **9 Supp 1**, 320 (2020).

21. Wells, G. A. *et al.* The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. (2000).

22. Mamluk, L. *et al.* Evidence of detrimental effects of prenatal alcohol exposure on offspring birthweight and neurodevelopment from a systematic review of quasi-experimental studies. *International Journal of Epidemiology* **49**, 1972–1995 (2020).

23. Saran, A. & White, H. Evidence and gap maps: A comparison of different approaches. *Campbell Systematic Reviews* **14**, 1–38 (2018).

24. McGuinness, L. A. & Higgins, J. P. T. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Research Synthesis Methods* **12**, 55–61 (2021).

25. Issue #102 - Feature request: Paired forest/risk-of-bias plot. *GitHub · mcguinlu/robvis*

26. Feinstein, A. R. & Cicchetti, D. V. High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology* **43**, 543–549 (1990).

27. Rosenthal, R. The file drawer problem and tolerance for null results. *Psychological Bulletin* **86**, 638–641 (1979).

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

29. Yamada, M. *et al.* Incidence and risks of dementia in Japanese women: Radiation Effects Research Foundation Adult Health Study. *Journal of the Neurological Sciences* **283**, 57–61 (2009).

30. Yamada, M. *et al.* Incidence and risks of dementia in Japanese women: The adult health study. *Journal of the Neurological Sciences* **283**, 306 (2009).

31. 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).

32. 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).

33. 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

34. 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).

35. Kuźma, E. *et al.* Which Risk Factors Causally Influence Dementia? A Systematic Review of Mendelian Randomization Studies. *Journal of Alzheimer's disease: JAD* **64**, 181–193 (2018).

36. Anstey, K. J., Ashby-Mitchell, K. & Peters, R. Updating the Evidence on the Association between Serum Cholesterol and Risk of Late-Life Dementia: Review and Meta-Analysis. *Journal of Alzheimer's Disease* **56**, 215–228 (2015).

37. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *bioRxiv* (2017). doi:10.1101/168674

38. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications* **9**, 224 (2018).

39. Sohani, Z. N. *et al.* Assessing the quality of published genetic association studies in meta-analyses: The quality of genetic studies (Q-Genie) tool. *BMC Genetics* **16**, 50 (2015).

40. Larsson, S. C. *et al.* Modifiable pathways in Alzheimer's disease: Mendelian randomisation analysis. *BMJ* **359**, j5375 (2017).

41. Waffenschmidt, S. *et al.* Development and validation of study filters for identifying controlled non-randomized studies in PubMed and Ovid MEDLINE. *Research Synthesis Methods* **11**, 617–626 (2020).

42. Wagner, M., Rosumeck, S., Küffmeier, C., Döring, K. & Euler, U. A validation study revealed differences in design and performance of MEDLINE search filters for qualitative research. *Journal of Clinical Epidemiology* **120**, 17–24 (2020).

43. Benn, M., Nordestgaard, B. G., Frikke-Schmidt, R. & Tybjærg-Hansen, A. Low LDL cholesterol, PCSK9 and HMGCR genetic variation, and risk of Alzheimer's disease and Parkinson's disease: Mendelian randomisation study. *BMJ* **357**, j1648 (2017).

44. Benn, M. Re: Low LDL cholesterol, PCSK9 and HMGCR genetic variation, and risk of Alzheimer's disease and Parkinson's disease: Mendelian randomisation study. *BMJ* (2017).

45. Booth, A. *et al.* An international registry of systematic-review protocols. *The Lancet* **377**, 108–109 (2011).

46. Danko, K. J., Dahabreh, I. J., Ivers, N. M., Moher, D. & Grimshaw, J. M. Contacting authors by telephone increased response proportions compared with emailing: Results of a randomized study. *Journal of Clinical Epidemiology* **115**, 150–159 (2019).

47. Meursinge Reynders, R., Ladu, L. & Di Girolamo, N. Contacting of authors modified crucial outcomes of systematic reviews but was poorly reported, not systematic, and produced conflicting results. *Journal of Clinical Epidemiology* **115**, 64–76 (2019).

48. Cooper, C., Bou, J. T. & Varley-Campbell, J. Evaluating the effectiveness, efficiency, cost and value of contacting study authors in a systematic review: A case study and worked example. *BMC Medical Research Methodology* **19**, 45 (2019).

49. Hennessy, E. A. *et al.* Ensuring Prevention Science Research is Synthesis-Ready for Immediate and Lasting Scientific Impact. *Prevention Science* (2021). doi:10.1007/s11121-021-01279-8

50. Hsieh, C.-Y. *et al.* Taiwan's National Health Insurance Research Database: Past and future. *Clinical Epidemiology* **11**, 349–358 (2019).

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

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

53. Jeyaraman, M. M. *et al.* Methodologically rigorous risk of bias tools for nonrandomized studies had low reliability and high evaluator burden. *Journal of Clinical Epidemiology* **128**, 140–147 (2020).