Lipids and dementia
An investigation of their relationship

Luke A McGuinness

University of Bristol

A thesis submitted for the degree of
Doctor of Philosophy in Population Health Sciences

TBC

Abstract

Background
In the UK, an estimated 800000 people are currently living with dementia and this number is expected to double by 2040. Despite the number of dementia cases and decades of research, there remains much unknown about the pathogenesis and progression of the disease, and, at present, no effective treatment exists to arrest or reverse the cognitive decline associated with the condition. In this context, identification of causal relationships between modifiable targets and dementia risk is central to the development of evidence-based prevention strategies and will be critically important in maintaining the long-term health of the ageing public. Blood lipid levels have been implicated in the aetiology of dementia by genetic linkage and functional cell biology studies, but current epidemiological evidence has yet to reach a consensus on their role in dementia risk.

Lipids and dementia
An investigation of their relationship

Luke A McGuinness

University of Bristol

A thesis submitted for the degree of

Doctor of Philosophy in Population Health Sciences

TBC

For Brendan McHugh

Acknowledgements

This is where you will normally thank your advisor, colleagues, family and friends, as well as funding and institutional support. In our case, we will give our praises to the people who developed the ideas and tools that allow us to push open science a little step forward by writing plain-text, transparent, and reproducible theses in R Markdown.

Luke McGuinness
Canynge Hall, Bristol
1 December 2021

Abstract

Background

In the UK, an estimated 800000 people are currently living with dementia and this number is expected to double by 2040. Despite the number of dementia cases and decades of research, there remains much unknown about the pathogenesis and progression of the disease, and, at present, no effective treatment exists to arrest or reverse the cognitive decline associated with the condition. In this context, identification of causal relationships between modifiable targets and dementia risk is central to the development of evidence-based prevention strategies and will be critically important in maintaining the long-term health of the ageing public. Blood lipid levels have been implicated in the aetiology of dementia by genetic linkage and functional cell biology studies, but current epidemiological evidence has yet to reach a consensus on their role in dementia risk.

Contents

List of Figures

List of Tables

List of Abbreviations

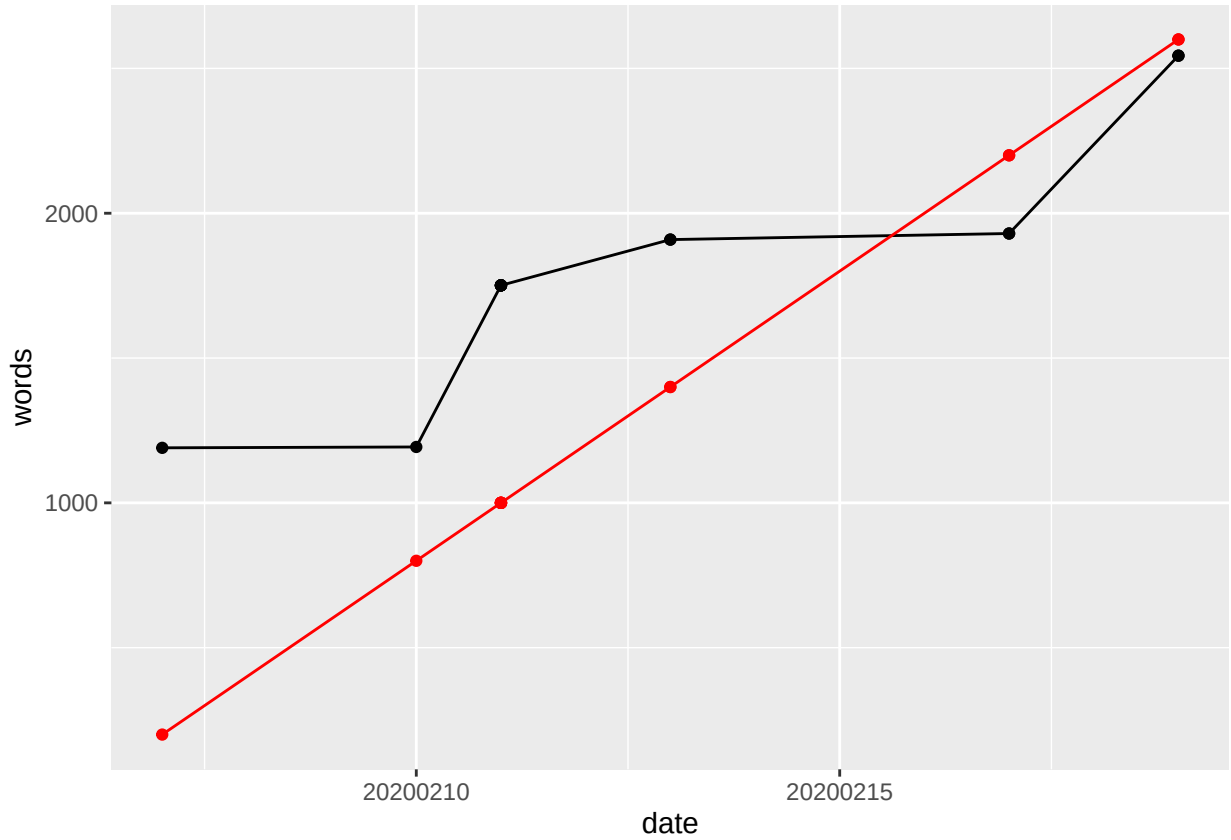
1-D, 2-D . . . One- or two-dimensional, referring in this thesis to spatial dimensions in an image.

Otter . . . . . One of the finest of water mammals.

Hedgehog . . . Quite a nice prickly friend.

Preface

- Word count: 2544 words
- Change since last day: 0 words
- Days: 13 days
- Words per day: 195.6923077



1 Introduction

Contents

1.1 Aims and Objectives of the thesis

1.1.1 Aim

1.1.2 Objective

1.2 Chapter Overview

- Chapter 4
- Chapter **??**
- Chapter **??**

1.3 Thesis Output

1.3.1 Peer reviewed papers

1.3.2 Papers under review

1.3.3 Software

- `robvis`: An R package and associted `shiny` web application that allows users to easily visualise the results of risk of bias graphs.
- `medrxvir`: An R package

1.3.4 Talks

1.4 Summary

-
-
-

2
Background

2.1 Dementia

2.1.1 History

2.1.2 Economic impact

2.1.3 Risk factors

2.1.4 Treatments

2.1.5 Preventative measures

2.2 Serum lipids

2.2.1 Range of lipids

2.2.2 HDL-c

2.2.3 LDL-c

2.2.4 Triglycerides

2.2.5 Total cholesterol

2.3 Summary

Why are open source statistical programming languages the best?

Because they R.

— Bealy, 2013[1]

3

Creating new systematic review tools in R

Contents

3.1 Background

R is a pro Below I introduce two R packages that I created to faciliate some aspects of the systematic review process.

3.2 robvis

3.2.1 Introduction

3.2.2 Installation

A stable version of `robvis` is hosted on the Comprehensive R Archive Network (CRAN) and can be installed using:

```
install.packages("robvis")
```

As development of `robvis` is ongoing, new features are often available in the development version some time before they appear in the stable CRAN version. The most recent development version can be install from GitHub using:

```
devtools::install_github("mcguinlu/robvis")
```

3.2.3 Reception and Future Plans

As of 19 February, 2020, `robvis` has been downloaded more than 1200 times. It has been accepted

It has also been intergrated with a suite of new online platforms for performing risk of bias assessments online.

In this thesis, `robvis` is used to present the results of the risk of bias assessments conducted as part of the systematic review and

While `robvis` currently provides a stable, future development work is planned.

3.3 medrxivr

3.3.1 Introduction

Searching pre-print repositories is becoming an increasingly important part of a systematic review. However, while several of the main repositories (e.g. arXiv, bioRxiv) have existing methods by which records can be searched and downloaded en masse, this is not true for medRxiv, an offspring of bioRxiv which hosts preprints in the medical, clinical, and related health sciences.

At present, medRxiv allows only simple search queries, as opposed to the often complex boolean logic that information specialists use to query other major databases. Additionally, record details must be downloaded individually, rather than in batches, making the export of relevant records for title and abstract screening a time consuming task.

Im order to facilitate the searching

3.3.2 Installation

To install `medrxivr` from CRAN:

```
# Currently doesn't work as not on CRAN


install.packages("medrxivr")
```

To install the development version from GitHub:

```
devtools::install_github("mcguinlu/medrxivr")
```

3.3.3 Methods

The `medrxivr` project is split into two parts:

- A webscraper that runs once a day which collects and cleans new records uploaded to medRxiv, and adds them to a machine readable database.
- A lightweight R package that provides an interface for this database, allowing users to search for relevant records, and easily retieve the associated full tex PDFs.

The webscraper is a straightforward R script built in R using `rvest`[cite] and a range of text processing packages[cite]. To quickly illustrate the process, the code to retrieve the total number of records on medRxiv is included below:

```
# Include and highly comment code.
```

Using Windows Task Scheduler, the script is scheduled to automatically run every morning, adding new records uploaded to medRxiv since the last run of

the webscraper to a local comma-seperated-values (CSV) file. Following quality control checks, the updated dataset is automatically uploaded to a cloud server, and immediately becomes available to the main `medrxivr` package functions.

The functions in the `medrxivr` package then facilitate users in working with this dataset. There are two main functions and a helper function:

- `mx_search()` [main]: Enables users to search the medrxivr data dump, using regular expressions and boolean logic.

```
topic1  <- c("dementia","vascular","alzheimer's")  # Combined with OR
topic2  <- c("lipids","statins","cholesterol")     # Combined with OR

myquery <- list(topic1, topic2)                    # Combined with AND

results <- mx_search(myquery)
```

- `mx_download()` [main]: Takes the output from `mx_search()` and retrieves the full text PDF for each record, saving it to a folder specified by the user.

```
mx_download(results,        # Object returned by mx_search
            "pdf/")         # Directory to save PDFs to
```

- `mx_crosscheck()` [helper]: Provides information on the version of the data dump that the user is searching, and checks whether any new records have been uploaded to medRxiv since the last run of the webscraper.

```
mx_crosscheck()
```

```
## Using medRxiv DataDump - 2020-02-19 10:38
## 2 new record(s) added to medRxiv since last data dump
```

3.3.4 Reception and Future Plans

`medrxivr`

Published paper,

3.4 Dicussion

3.5 Summary

*In this Chapter, I have introduced two new tools for facilitating evidence syntheses in R: `robvis` and `medrxivr`

- I have
- The impact of these packages to date, and a roadmap for their future development, has been discussed.

Hold

— Charles Darwin[2]

4

Chapter 2: Systematic Review

Contents

4.1 Additional ideas

- Evidence map - show the distribution of the different studies population across the world.
- Living systematic review approach - update weekly based on medRxiv

4.2 Aims

The aim of this chapter is to systematically review all available literature on the association between blood levels of total cholesterol and it's constituent parts (HDL-c,LDL-c and triglycerides) on the subsequent risk of dementia.

4.3 Methods

4.3.1 Search strategy

We will systematically search electronic bibliographic databases to identify potentially relevant records. The search strategy used in each database will be developed in an iterative manner using a combination of free text and controlled vocabulary (MeSH/EMTREE) terms to identify studies which have examined the relationship between blood lipids levels and dementia, incorporating input from an information specialist. The strategy will include terms related to lipids, lipid modifying treatments, and dementia and its sub-types, and will be 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 3.2 below and the full draft search strategies for each database are attached to this protocol. 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 Line 7 NOT Line 13 in Table 3.2) will be screened. If any potentially relevant studies are identified, their titles and abstracts will be searched for key terms that can be incorporated into the filters to improve search sensitivity.

The following databases will be searched from inception onwards: Medline, EMBASE, Psychinfo, Cochrane Central Register of Controlled Trials (CENTRAL), and Web of Science Core Collection. We will also search clinical trial registries, for example ClinicalTrials.gov, to identify relevant randomized controlled trials.

The abstracts list of relevant conferences (e.g. the proceedings of the Alzheimer's Association International Conference, published in the journal Alzheimer's &

Dementia) will be searched. Grey literature will also be searched via ProQuest, OpenGrey and Web of Science Conference Proceedings Citation Index, while theses will be accessed using the Open Access Theses and Dissertations portal. We will also search BioArxiv, a preprint repository, to identify potentially relevant studies. Finally, the reference lists of included studies will be searched by hand while studies citing included studies will be examined using Google Scholar (forward and reverse citation searching).

4.3.2 Study selection

Records will be imported into Endnote and deduplicated using the method outlined in Bramer et al. (2016).[3] Screening (both title/abstract and full-text) will be performed using a combination of Endnote and Rayyan, a web based screening application.[4] Title and abstract screening to remove obviously irrelevant records will be performed by the primary author, with a random selection of excluded records being screened in duplicate to ensure consistency with the inclusion criteria. If this demonstrates a significant level of erroneous exclusion by the primary author a larger proportion will be dual-screened. Full-text screening will also be completed in full by the primary author. A second reviewer will screen a random sample of included and excluded records, in addition to any records identified by the first reviewer as being difficult to assess against the inclusion criteria. Reasons for exclusion at this stage will be recorded. Disagreements occurring during either stage of the screening process will be resolved through discussion with a senior colleague. A PRIMSA flow diagram will be produced to document how records moved through the review.[5]

Inclusion

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

Eligible studies will define dementia according to recognised criteria, for example the National Institute of Neurological Disorders and Stroke Association-Internationale pour la Recherche en l'Enseignement en Neurosciences (NINDS-AIREN), International Classification of Diseases (ICD), or Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria. For MCI, eligible studies are those that attempted state a definition for diagnoses of MCI (e.g. an adapted version of the Petersen criteria[6]) and create ordinal groups of patients (e.g. no dementia or dementia/MCI/dementia) based on this definition.

No limitations will be imposed on publication status, publication date, venue or language, although we will require sufficiently detailed reports of the studies to be able to examine their methods. Preprints and unpublished reports will be eligible for inclusion if relevant. Multiple publications resulting from the analysis of the same data will be included and grouped.

Exclusion

Case-control studies, cross-sectional studies, qualitative studies, case reports/series and narrative reviews will be excluded. Studies which present no evidence of attempting to exclude prevalent cases from their analyses will also be 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) will be excluded. Conference abstracts with no corresponding full-text publication will be examined, and we will contact 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) will be excluded. Previous systematic

reviews will not be eligible, but their reference lists will be 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 "multidomain 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) will be excluded.

4.3.3 Data extraction

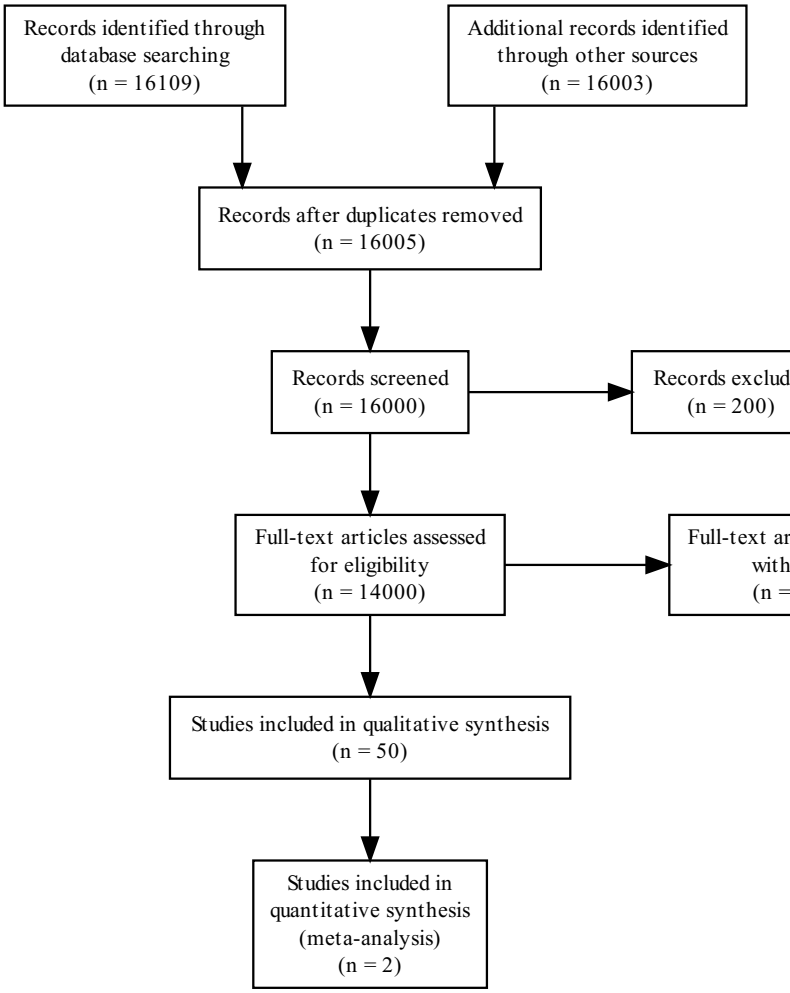
4.3.4 Risk of bias assessment

Risk of bias assessment was performed using the domain-based risk-of-bias assessment tool appropriate to the study design. Randomised controlled trials were assessed using the RoB2 tool,[7], non-randomised studies of interventions were assessed using the ROBINS-I tool,[8] and non-randomised studies of exposures were assessed using the ROBINS-E tool.[TBC]

At present, no risk of bias assessment tool for Mendelian randomisation studies is available. Bias in thesee studies was assessed with the help of an expert panel [TBC]

4.3.5 Patient and public involvement

4.4 Results



4.5 Discussion

4.6 Conclusion

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

— Harold F. Dorn, 1953[9]

5

Chapter 3: Primary analysis of lipid regulating agents and dementia

Contents

5.1 Additional ideas

One of the problems I'll point out with the trials in the systematic review chapter (Chapter 4) is that they only included people with a high cardiovascular risk - we are kind of doing the same by using elevated test cholesterol results as the index event.

5.2 Aims

5.3 Methods

5.3.1 Estimation methods

Potential biases included time varying confounding, selection bias due to censoring on death and We performed a Cox

5.3.2 Estimating the value of the time-varying confounders

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

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

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

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

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

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

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

i.e.

Timepoint 12345678

CVD 00001111

Treatment 00000111

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

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

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

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

However, this is not the case. Baseline cholesterol level are predicted to be a poorer instrument for than QRISK2 score,[10] which estimates a patients' 10-year risk of a cardiovascular event. Current NICE guidelines state that those with a QRISK score of 10% or higher, and in whom lifestyle modifcation is ineffective/inappropriate, should recieved a lipid regulation agent. However, this analysis could not find any effect of QRISK2 scores on statins precription levels at 6 months. [Need to cite Lauren's eventual paper here focusing on QRISK2, but also display a RD analysis of TC/LDL-c levels here on statins at 6 months. Need also to check, as Lauren mentioned she found some evidence that there is a relationship in practices that acutally did what they should.]

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

6
References

7
Annotated Bibliography

Appendices

A
By Chapter

A.1 Chapter 1

A.2 Chapter 2

B

Other Appendix

1. Bealy, C. R Programming Humour. *Cross Validated* (2013).

2. Chang, C.-C. H., Zhao, Y., Lee, C.-W. & Ganguli, M. Smoking, death, and Alzheimer's disease: A case of competing risks. *Alzheimer Disease and Associated Disorders* **26**, 300–306 (2012).

3. 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).

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

5. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration | The BMJ.

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

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

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

9. Dorn, H. F. Philosophy of Inferences from Retrospective Studies. *American Journal of Public Health and the Nations Health* **43**, 677–683 (1953).

10. Hippisley-Cox, J. *et al.* Predicting cardiovascular risk in England and Wales: Prospective derivation and validation of QRISK2. *BMJ* **336**, 1475–1482 (2008).