1

Individual participant data meta-analysis

1.1 Lay Summary

As part of a broader investigation into the relationship between lipid levels and dementia risk, I sought to

1.2 Introduction

Is there a good way to visualise input of cohorts into an IPD analysis?

Justify analysis based on findings of systematic review - suggestion of variation of effect by age, specifically

1.3 Methods

1.3.1 Applying for data access

Systematic review

As part

As part of the IPD analysis, relevant cohort studies identified through the systematic review detailed in Chapters **??** were. In the first instance, the corresponding author on the main man

Dementia platform UK

In addition, to assess the results for the Dementia Platform UK common-access procedure was utilised. In short, this approach is intended to make it easier to access data from exisiting dementia cohorts through a centralised application process.

Talk about how different covariates were managed across cohorts,

1.3.2 Eligibility criteria

Lipids reported/available as a continuous measure, and were not cross-sectional.

Due to the limited time-frame, studies making use of population-level electronic health records, which often require an entire project proposal, were ineligible due to the time and cost involved in applying. The one exception to this was data from the CPRD, which we already had access to via the study reported in Chapter 'ref(cprd-heading).

1.3.3 Missing data

Talk about missing data were handled

1.3.4 Risk of bias assessment

Risk of bias assessment was performed for each of the included cohorts using the relevant tool described in Section **??**.

There is some concerns about performing risk of bias assessments on your own analysis, and so

1.3.5 Analysis

All analysis were standardised by changes in 1-SD of the exposure variable. This was done to aid interpretability of the outcome.

Similary,

Clustering within studies was accounted for, given the evidence that ... [@abo-zaid2013]

Stratified by age-at-entry and sex, and ethnicity where possible.

Hazard ratios were utilised to

In order to investigate the interaction of sex with

A two-stage model was used out of necessity, with the one-stage model being precluded by the different datasets being in protected data silos.

A discrete proportional hazards model was employed to account for the interval censoring introduced by design o fthe longitudinal cohort studies.[@wang2017]

In order to investigate the interaction of patient-level characteristics with lipid levels, interaction terms for lipid-covariate terms were included in the model above. These were extracted and synthesised using a random effects meta-analysis.

Where results were comparably to previously published estimates, these were compared and reasons for any discrepancies discussed.

The analysis were carried out in R (Version...) using the following packages: `survival`[@survival-book], `metafor`[@viechtbauer2010a]

Replicating published findings

Where any included data source had previously been analysed and results for the association between lipid levels and dementia reported, these were compared to the results of this analysis.

1.4 Results

1.4.1 Data access

Of the XXX studies to which I applied for data access, only three were eventually included in the analysis. Figure 1.1 details whether the cohorts eventually included in the review were identified by the systematic review or via DPUK portal.

In addition, the reasons for cohorts not being included in the analysis are presented. In summary, the requests for data from cohorts identified by the systematic review was characterised by a very low response rate. For the minority who did respond, common reasons given by corresponding authors for not sharing the data included that they no longer worked in the lab, had access to the data, or that they were currently or intended to perform a similar analysis as the one proposed.

For the streamlined DPUK process, where a dedicated project manager liases with cohort owners, the overall response rate was higher, result in access to three cohorts. However, even using th

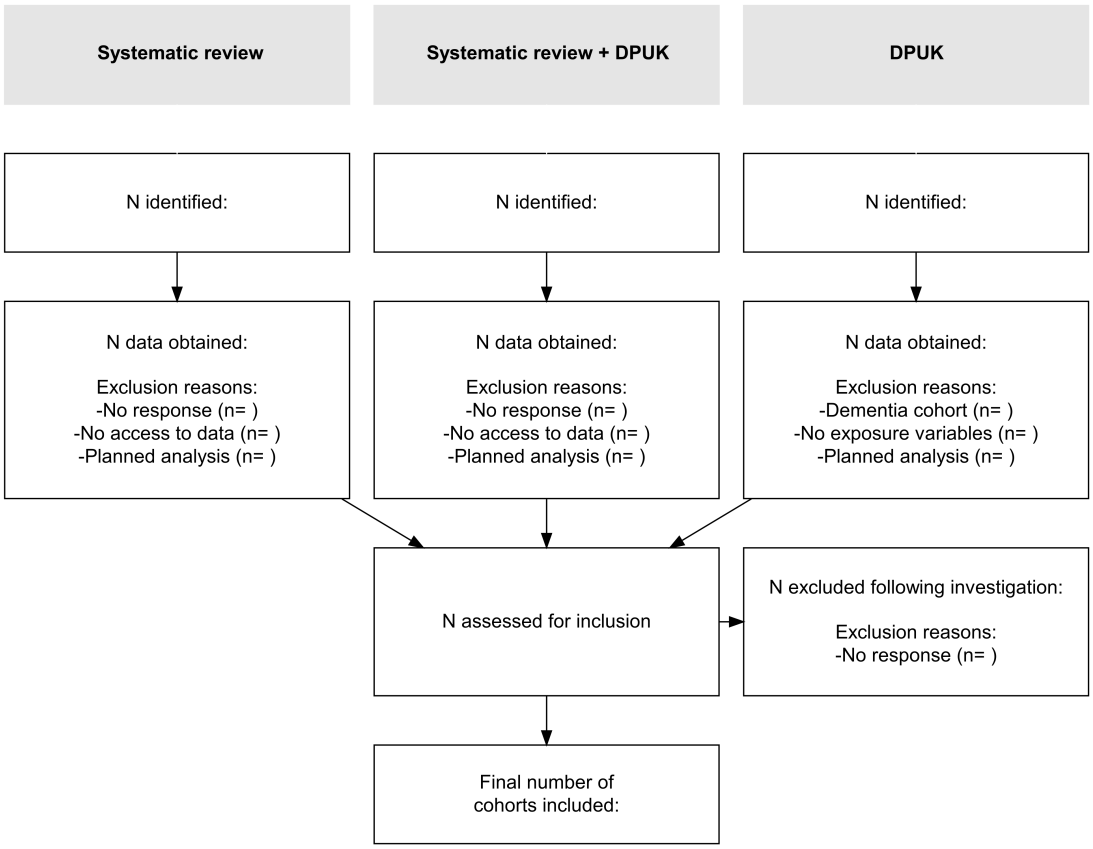

Figure 1.1: Flowchart of included cohorts, stratified by identification method (systematic review vs DPUK).

Few cohorts were included in both the DPUK and the systematic review sets of cohorts indicating that the DPUK.

I did not request data from several DPUK cohorts due to the information on the online system indicating that the study was relatively new (and so was yet to collect >1 wave of data, as in the case of the XXXX cohort) or the online data dictionary indicating that the exposure variables of interest were not recorded.

As highlighted in the

1.4.2 Included data sources

The three datasets included are described in detail in the following section, and are summarised in Table 1.1.

Of note, all data sources included in the analysis were based in the United Kingdom. This is likely due to the majority of included datasets coming from the Dementia Platform UK route 1.1, which as implied by the name, has a geographical focus on studies performed in Great Britain United Kingdom.

Table 1.1: Summary of cohorts for which data were available

Cohort	**N**	**Dementia events (all-cause)**	**Age (mean)**	**Male (%)**
CaPS	2512	1034	52	100
CPRD	X	X	X	X
EPIC	1001	5	52	45
Whitehall II	8022	181	50	69

Caerphilly Prospective Study

The Caerphilly Prospective Study is a longitudinal study of men in the

Cholesterol measures (total, LDL-c, HDL-c and triglycerides) were measured at baseline in 1979-1983. As the study population has aged, additional outomes. Of particular relevance to this analysis, from Phase III (1989-1993) onwards, a battery of congnitive tests were introduced.

CPRD

The Clinical Practice Research Datalink (CPRD) is a large population-based, electronic health record (EHR) database.[@herrett2015] containing the primary care records for more than 10 million primary care patients in England, and is

broadly representative of the UK population in terms of age, sex and ethnicity.[@herrett2015; @mathur2014]

The CPRD is introduced more fully in Chapter @ref(). Briefly, a similar approach to the cohort definition as used in Chapter

Participants were included from the first date of lipid measurement, so no issues with immortal time bias as discussed in Chapter ...

Additionally, ther number of participants is larger in this analysis as there is no restriction on the level which lipids should be in order to be included in the analysis.

the

Epic Norfolk

The European Prospective Investigation of Cancer - Norfolk is a [@riboli1997; @riboli2002]

Has only 8 events

Whitehall II

The Whitehall II study is a prospective cohort study of 10 308 participants (70% men), aged 35–55 years and recruited between 1985 and 1989 from 20 London-based Civil service departments (`https://www.ucl.ac.uk/whitehallII`). Clinical examinations have been performed in 1991-1994, 1997-1999, 2002-2004, 2007-2009, 2012-2013, 2015-2016 with the data from circulating metabolomic traits and cognitive testing for the present study obtained from the 1997-1999 clinic phase. - taken from EN ID: 2140

This data source was analysed in one of the included studies identified by the systematic review presented in Chapter **??**.[@tynkkynen2018]

This gave

1.4.3 Cohorts provided data but ultimately excluded

As highlighted in Figure 1.1, several cohorts from the DPUK responding positively but on inspection of data provided these cohorts were excluded.

Several cohorts were excluded on the basis of a lack of exposure variables, including Cam-CAN (had cardiovascular category, but only contained blood pressure).

The reasons for exclusion of these cohorts for which data were provided is illustrated in Table 1.2.

BRACE

Likley had dementia at baseline

Memento

Excluded as criteria for entry was outpatients from memory clinic (unlikely to be dementia free at baseline)

Generation Scotland

Cross-sectional data only

NICOLA

NICOLA study - only had cross-sectional data

ELSA

Data provided but only had ever high cholesterol as a binary variable. Not compatible with

```
## Warning in read.table(file = file, header = header, sep = sep, quote = quote, :
## incomplete final line found by readTableHeader on 'C:/Users/lm16564/OneDrive -
## University of Bristol/Documents/rrr/thesis/data/ipd/dataExcluded.csv'
```

Table 1.2: dataExcluded

Cohort	Reason
NICOLA	Cross-sectional - only one wave of data available
TRACK HD	Participants carried HTT gene (i.e. premanifest Huntington's Disease). Cohort owners indicated that this is likely to overshadw

1.4.4 Cohorts identified the systematic review but excluded from the IPD analysis

Electronic health record co

LBC - 1936

Taiwan etc,

1.4.5 Cohorts approach but received no answer

A particularly frustrating in the case of cohorts which had a dedicated data access panel, for example, the Three City Study. Despite multiple attempts to contact the team, there was no response received.

1.4.6 Data cleaning and harmonisation

Across all cohorts, data cleaning was performed in a similar manner, using commonly named variables, so that a single model could be applied using functional programming.

The one exception to this is the CPRD data, which was held in a different system to the rest.

The advantage of this approach is that it reduces the liklihood of errors in model mis-speciication if needing to change variables names from cohort to cohort.

For all cohorts, the first lipid measurement was used for the exposure method,

For heterogeneity across the cohorts, the total time-at-risk was investigated.

1.5 Discussion

Useful citation for discussion[@levis2021]

1.5.1 Limitations

Low response rate to request for data

This review had a low response rate to requests for data access.

While this is not unexpected, given that a review of IPD studies published between 1987 and 2015 found that fewer than half managed to obtained data from greater than 80% of studies, and that in many cases, the exact percentage of studies for which data were obtained was not accurately reported. test There are likely several reasons for this.

- Individual participant data meta-analysis including studies other than randomised controlled trials have less success in obtaining individual participant data from studies.[@nevitt2017a]
- While there is no

For many other reasons also - if

A potential further reason was highlighted during my attendance at the

Letters sent to all cohorts identified through the systematic review

This is likely due to my junior position as a early career research combined with the impact

Range of reasons why data is not readily available. Privacy concerns, concerns around scoping or "parasitic" behaviour, and a lack of trust (i.e. primary researchers do not trust secondary researchers).

Unfortunately,

The low response rate and the

Whether or not to press ahead with an IPD analysis in the absence of all (or even most) data is a personal decision, and some have highlighted where they decided not to pursue an IPD study.[@jaspers2014]

For the purposes of this thesis, conducting the IPD was useful as it afforded me the opportunity to experience the methodology.

On a personal level, given the

In addition, it allowed me to analysed two addition previously unanalysed cohort studies which were then utilised in the triangulation study detailed in Chapter **??**.

1.5.2 Resources

While it was disappointing that such a small proportion of the available data were obtained for analysis, in terms of project completion, it is probably a good thing.

Given the resource intensity of data cleaning and harmonisation,

1.5.3 Data access

Couple of sections

- Problems with getting data from studies

-

Sharing data has several legal, ethical and logistical challenges, and can often

This is in theory what the DPUK was built to address. However, despite claims to a streamlined process, the response rate amond cohorts participating

Describe your experience of trying to access the DPUK - while a great resources, frustrating at times. The portal would benefit from a centralised management system and

Even with the guaranteed data access afforded by the DPUK, accessing sufficient data were

Of the XXX cohorts requested from the DPUK, allowing for a year lag since application, a response was received from only XXX (XXX%), while of those responding a minority

1.5.4 Study within a review

In future, it could have been worth running a "study within a review", such as that used to examine the best method [@godolphin2019a] to identify and access study data.

Options could include

Previous SWARs have been registered as protocols, but have not yet reported!

1.5.5 Reflections on the process

May have been overly ambitious

1.6 References