Medicines & Healthcare products Regulatory Agency

CPRD

INDEPENDENT SCIENTIFIC ADVISORY COMMITTEE (ISAC) PROTOCOL APPLICATION FORM

PART 1: APPLICATION FORM

IMPORTANT

Both parts of this application must be completed in accordance with the guidance note 'Completion of the ISAC Protocol Application Form', which can be found on the CPRD website cprd.com/research-applications

FOR ISAC USE ONLY	
Protocol No. -	**Submission date -**

GENERAL INFORMATION ABOUT THE PROPOSED RESEARCH STUDY

1. Study Title (Max. 255 characters)

The effect of common infections on the risk of dementia in individuals with and without diabetes: a cohort study using UK primary and secondary care data

2. Research Area (place '**X**' in all boxes that apply)

Drug Safety		Economics	
Drug Utilisation		Pharmacoeconomics	
Drug Effectiveness		Pharmacoepidemiology	
Disease Epidemiology	X	Methodological	
Health Services Delivery			

3. Chief Investigator

Title:	Dr
Full name:	Charlotte Warren-Gash
Job title:	Associate Professor of Epidemiology/ Wellcome Intermediate Clinical Fellow
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	
CV Number (if applicable):	

4. Corresponding Applicant

Title:	Ms
Full name:	Rutendo Muzambi
Job title:	PhD Student
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	
CV Number (if applicable):	

Medicines & Healthcare products Regulatory Agency

CPRD

5. List of all investigators/collaborators

Title:	Ms
Full name:	Rutendo Muzambi
Job title:	PhD Student
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	
CV Number (if applicable):	
Will this person be analysing the data? (Y/N)	Y

Title:	Professor
Full name:	Liam Smeeth
Job title:	Professor of Clinical Epidemiology
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	
CV Number (if applicable):	
Will this person be analysing the data? (Y/N)	N

Title:	Professor
Full name:	Krishnan Bhaskaran
Job title:	Professor in statistical epidemiology and Sir Henry Dale fellow
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	
CV Number (if applicable):	
Will this person be analysing the data? (Y/N)	N

Title:	Professor
Full name:	Carol Brayne
Job title:	Director of the Cambridge Institute of Public Health
Affiliation/organisation:	Cambridge University
Email address:	
CV Number (if applicable):	
Will this person be analysing the data? (Y/N)	N

Title:	Professor
Full name:	Nish Chaturvedi
Job title:	Professor of Clinical Epidemiology
Affiliation/organisation:	University College London
Email address:	
CV Number (if applicable):	
Will this person be analysing the data? (Y/N)	N

6. Experience/expertise available

List below the member(s) of the research team who have experience with CPRD data.

Name:	**Protocol Number/s:**
Dr Charlotte Warren-Gash	17_176R, 18_134R, 19_096
Professor Liam Smeeth	12_027RA, 12_065, 15_257, 16_174, 16_113A, 18_207, 18_278
Professor Krishnan Bhaskaran	12_090, 10_097, 12_044, 12_027, 16_174, 16_113A

List below the member(s) of the research team who have statistical expertise.

Name(s):	

Professor Krishnan Bhaskaran

List below the member(s) of the research team who have experience of handling large datasets (greater than 1 million records).

Name(s):
Professor Krishnan Bhaskaran
Professor Liam Smeeth

List below the member(s) of the research team, or supporting the research team, who have experience of practicing in UK primary care.

Name(s):
Ms Rutendo Muzambi
Professor Liam Smeeth

ACCESS TO THE DATA

7. Sponsor of the study

Institution/Organisation:	London School of Hygiene and Tropical Medicine
Address:	Keppel Street, London, WC1E 7HT

8. Funding source for the study

Same as Sponsor?	Yes		No	X
Institution/Organisation:	Alzheimer's Society			
Address:	Alzheimer's Society, 43-44 Crutched Friars, London, EC3N 2AE			

9. Institution conducting the research

Same as Sponsor?	Yes	X	No	
Institution/Organisation:	London School of Hygiene and Tropical Medicine			
Address:	Keppel Street, London, WC1E 7HT			

10. Data Access Arrangements

Indicate with an '**X**' the method that will be used to access the data for this study:

Study-specific Dataset Agreement	

Institutional Multi-study Licence	X
Institution Name	London School of Hygiene and Tropical Medicine
Institution Address	Keppel Street, London, WC1E 7HT

Will the dataset be extracted by CPRD?

Yes		No	X

If yes, provide the reference number:

11. Data Processor(s):

Processing	X
Accessing	X
Storing	X

Processing area (UK/EEA/Worldwide)	UK
Organisation name	London School of Hygiene and Tropical Medicine
Organisation address	Keppel Street, London, WC1E 7HT

[Add more processors as necessary by copy and pasting a new table for each processor]

INFORMATION ON DATA

12. Primary care data (place '**X**' in all boxes that apply)

CPRD GOLD	X	CPRD Aurum	

13. Please select any linked data or data products being requested

Patient Level Data (place '**X**' in all boxes that apply)

ONS Death Registration Data		CPRD Mother Baby Link	
HES Admitted Patient Care	X	Pregnancy Register	
HES Outpatient		NCRAS (National Cancer Registration and Analysis Service) Cancer Registration Data	
HES Accident and Emergency		NCRAS Cancer Patient Experience Survey (CPES) data	
HES Diagnostic Imaging Dataset		NCRAS Systemic Anti-Cancer Treatment (SACT) data	
HES PROMS (Patient Reported Outcomes Measure)		NCRAS National Radiotherapy Dataset (RTDS) data	
		Mental Health Services Data Set (MHDS)	

Area Level Data (place '**X**' in all boxes that apply)

Practice level (UK)		**Patient level (England only)**	
Practice Level Index of Multiple Deprivation (Standard)	X	Patient Level Index of Multiple Deprivation	X
Practice Level Index of Multiple Deprivation (Non-standard)		Patient Level Townsend Score	
Practice Level Index of Multiple Deprivation Domains (Non-standard)			
Practice Level Carstairs Index for 2011 Census (Excluding Northern Ireland) (Standard)			
2011 Rural-Urban Classification at LSOA level (Non-standard)			

Reference number (where applicable):

14. Are you requesting linkage to a dataset not listed above?

Yes		No	**X**

If yes, provide the reference number:

15. Does any person named in this application already have access to any of these data in a patient identifiable form, or associated with an identifiable patient index?

Yes		No	**X**

If yes, provide further details:

VALIDATION/VERIFICATION

16. Does this protocol describe an observational study using purely CPRD data?

Yes	**X**	No	

17. Does this protocol involve requesting any additional information from GPs, or contact with patients?

Yes		No	**X**

If yes, provide the reference number:

PART 2: PROTOCOL INFORMATION

Applicants must complete all sections listed below **Sections which do not apply should be completed as '*Not Applicable*' and justification provided**
A. Study Title (Max. 255 characters) The effect of common infections on the risk of dementia in individuals with and without diabetes: A cohort study using UK primary and secondary care data
B. Lay Summary (Max. 250 words) Dementia is a major public health burden posing a devastating impact on individuals, caregivers and healthcare services. In the UK, it was estimated that around 850,000 people were living with dementia in 2015 and this number is projected to rise to over 1 million by 2025. Due to the rising ageing population and lack of medications that can cure or prevent dementia, it has become increasingly important to identify factors that can reduce the risk of dementia. Over, the last few decades, there has been growing interest on the role of infections on the risk of dementia. However, it remains unclear whether people with common infections such as pneumonia and urinary tract infections, have a higher chance of developing dementia. Additionally, common infections frequently occur in people with diabetes, and in turn diabetes is associated with dementia. Therefore, we aim to carry out a study where we will follow individuals over time using anonymous data from primary care and hospital health records to investigate whether people with common infections are at an increased risk of developing dementia and whether this risk differs in people with and without diabetes. Infections and diabetes are potentially preventable and therefore a better understanding of how these conditions affect the risk of dementia could lead to important public health interventions. These interventions may include strategies to increase the uptake of vaccines to prevent infections, and early recognition and treatment of infections in people with diabetes to reduce the risk of developing dementia.
C. Technical Summary (Max. 300 words) Dementia poses a significant burden on disability and dependence worldwide. Due to the increasing ageing population and absence of pharmacological therapies that can delay the onset or progression of dementia, dementia risk reduction has become a public health priority. Recent evidence suggests that the incidence of dementia is declining in Europe and the USA, and this change has been partly attributed to modifiable risk factors. Common infections have been identified as potential risk factors for dementia. In turn, common infections are more prevalent in diabetes, which is a strong risk factor for dementia. We hypothesise that individuals diagnosed with common infections (lower respiratory tract, urinary tract, skin and soft tissue infections and sepsis) will have an increased risk of dementia, and that this risk will increase in individuals with diabetes compared to those without diabetes. To test this hypothesis, we will carry out a cohort study of older adults aged 65 years and over using prospectively collected CPRD data linked to hospital episode statistics. We will exclude individuals with prevalent dementia and cognitive impairment at baseline. We will assess the age-specific incidence rates of dementia in individuals with and without common infections. Then, we will use Cox regression models to investigate the effect of the type, timing and frequency of infections on the incidence of dementia, adjusting for confounding factors. We will then investigate the presence of effect modification by diabetes on the association between common infections and incident dementia. Finally, we will investigate whether there is an association between common infections and evidence of cognitive impairment. To the test the robustness of our findings, we will carry out a range of sensitivity analyses. Improved understanding of the interrelationship between infections and diabetes with incident dementia will help to inform dementia risk reduction interventions.

D. Outcomes to be Measured

Primary Outcomes

(1) Incidence of dementia (all types)
(2) Incident dementia (type-specific Alzheimer's disease, vascular, mixed, other).

Secondary Outcome

(1) Evidence of cognitive impairment

Objectives, Specific Aims and Rationale

Objective

The objective of this study is to investigate the effect of common infections (lower respiratory tract, urinary tract and skin and soft tissue infections and sepsis) on the incidence of dementia in adults aged 65 years and older and whether this risk varies in individuals with and without diabetes, using CPRD data linked to HES.

Specific aims

1. To describe the age-specific incidence rates of dementia in adults aged 65 years and older with and without common infections.
2. To investigate whether the presence, frequency, timing and type of common infections affect the risk of dementia.
3. To investigate whether diabetes modifies any association between common infections and incident dementia.
4. To investigate whether the presence or type of common infections are associated with evidence of cognitive impairment (secondary outcome).

Rationale

Identifying modifiable risk factors for dementia has become increasingly important given the increasing burden of dementia. As a result, dementia risk reduction is a public and global health priority. If diabetes and common infections interact to increase the risk of developing dementia, these potentially preventable conditions could be a target to reduce the risk of dementia.

E. Study Background

Dementia is a major public health challenge. With the global prevalence projected to rise from 47.5 million in 2015 to 135.5 million by 2050, the burden of dementia on individuals, caregivers and healthcare services is set to rise markedly [1]. Currently, there are no pharmacological therapies that can delay the onset or progression of dementia and as the ageing population continues to rise, dementia risk reduction has become a public health priority [2].

Recently, a large multi-area, population-based study from the UK reported a 20% decrease in the age-specific incidence of dementia in adults aged 65 years and older [3]. Other population based studies from Europe and the US have reported a declining trend in the age-specific incidence of dementia among older adults, [4-9] with improvements in education and vascular risk factors partly accounting for this change [10]. Therefore, identification of risk factors for dementia could play an important role in risk reduction. Although the single biggest predictor of dementia is age[11], a non-modifiable risk factor, population-based cohort studies have shown that addressing modifiable risk factors can reduce the risk of dementia by a third [12-14]. These risk factors include smoking, hypertension, education, socioeconomic status and diabetes.

Over the last few decades, there has been a large body of evidence to suggest that infections play a role in the risk of dementia. Pathological evidence has demonstrated the prevalence of bacterial, viral and fungal infections in individuals with Alzheimer's disease [15-17]. However, due to the cross-sectional nature of these studies, the ability to assess temporality or to make any inferences about causality is limited. Acute infections are well known to precipitate short term changes in cognition. However, the association of these infections with long term cognitive impairment is less established. Findings from a US prospective study of older adults showed that individuals hospitalised with sepsis, an acute life threatening infection, were likely to develop moderate to severe cognitive impairment [18]. In turn, cognitive impairment is a strong predictor for dementia [19].

Common infections such as pneumonia and urinary tract infections, have been shown to be prevalent in individuals with dementia [20, 21]. Few longitudinal studies have investigated the role of these infections on the incidence of dementia. Two of these studies were insufficiently powered and focused only on patients hospitalised with pneumonia, limiting their ability to capture patients with less severe infections [22] [23]. A large scale retrospective study of US veterans (N=417,172) found that the incidence of dementia was increased by the following infections: pneumonia (HR 1.10 95% CI 1.02-1.19), urinary tract infections (HR 1.13 95% CI 1.08-1.18), cellulitis (HR 1.14 95% CI 1.09-1.20) and sepsis [24]. However, the generalisability of the findings was restricted to males and military veterans, and as with the two aforementioned studies, the study was conducted in the US, limiting generalisability to other countries. Recently, a large-scale population-based study of over 60,000 individuals using a Taiwanese longitudinal health insurance database showed that individuals with a history of sepsis were at a greater risk of developing dementia compared to those without sepsis. However, the majority of these studies did not investigate a range of common infections within the same study and none of these studies examined the effects of multiple episodes of infection on the incidence of dementia. Additionally, although a longitudinal study using UK primary care data found that episodes of infection were associated with an increased likelihood of a diagnosis of dementia in elderly adults aged 84 years or older [25], no studies in the UK have specifically investigated the association between types and frequency of common infections with incident dementia.

Infections frequently occur in people with diabetes [26]. In a recent systematic review of cohort and case control studies, diabetes was associated with an increased incidence of infections including respiratory, genitourinary and skin infections [27]. Additionally, two studies using data from UK electronic health records showed that individuals with diabetes were at an increased risk of infections compared to the general population [28] [29]. Diabetes is a well-known risk factor for dementia. In a recent systematic review and meta-analysis of 14 prospective studies from 2.3 million people, diabetes was associated with a 60% increased risk of dementia overall [30]. These findings were consistent with two previous meta-analyses of longitudinal studies [31] [32]. Given the co-occurrence of infections and diabetes, and their association with dementia, it is possible that diabetes could modify the effect of common infections on the risk of dementia.

To our knowledge, no studies have investigated the potential interaction between diabetes and infections on the incidence of dementia. Diabetes and infections are both major health conditions that pose a significant impact on public health services. As both conditions are potentially preventable, understanding their association with dementia could have public health implications in targeting populations at an increased risk of dementia and early treatment of infections and diabetes could reduce the risk and burden of dementia.

Therefore, our aim is to investigate the effect of common infections on the risk of dementia and to investigate whether this effect varies in individuals with and without diabetes, using a large dataset of primary care records linked to hospital episode statistics (HES), representative of the UK population. We will also investigate the association between common infections and evidence of cognitive impairment.

Study Type

Descriptive

Hypothesis-testing

F. Study Design

Historical cohort study using CPRD data and linked HES data

G. Feasibility counts

- There were 1,009, 629 individuals aged 65 years and older in CPRD with linked HES data between 2004 and 2018 with at least 12 months research standard follow up and no prior history of dementia. 16.4% had any common infection.
- 11.4% had a lower respiratory tract infection, 0.7% had sepsis, 3.1% had a urinary tract infection and 4.2% had a skin and soft tissue infection. The median follow-up time was 12.7 years. 82% had more than 5 years of follow up.
- 4.7% (n=47589) developed dementia.
- The incidence of dementia in adults aged 65 years and older in the UK has been estimated to be around 209,600 new cases of dementia per year [3]. As dementia is underdiagnosed in primary care, it is likely that the incidence of dementia will be lower in CPRD.

H. Sample size considerations

We used the results of our feasibility counts to carry out our sample size calculations. Here we calculated the minimum effect estimate for common infections on the risk of dementia. Our estimates are conservative as we estimated that 16.4% of our study population had a first ever common infection and 4.7% of our total population developed dementia.

From our feasibility counts, we estimated that we would have 5 people unexposed to infections for every person with an infection. Hence, we will have an 80% power at a 5% significance level to detect a minimum hazard ratio of 1.02.

I. Planned use of linked data (if applicable):

We plan to use primary care data from CPRD linked to Hospital Episode Statistics. Although dementia cases are likely to be diagnosed in primary care settings, using HES will help to identify additional cases and improve the accuracy of information available on timing of dementia diagnosis. We will also identify infections using linked HES data: a recent UK study comparing incidence of community acquired pneumonia in primary and secondary care data among adults aged 65 years and older found that the incidence estimates of community acquired pneumonia were 28% lower in primary care data alone compared to linked data [33].

We also plan to use patient-level IMD and practice-level IMD as a measure of socioeconomic deprivation, which is a potential confounding factor. Our primary analysis will include only individuals with linked data, and therefore we plan to use patient-level IMD for this analysis. We will consider practice-level IMD for the sensitivity analysis which includes individuals without linked data.

J. Definition of the Study population

We will include all adults aged 65 years and older present in CPRD (Gold) with linked HES data, who were registered in CPRD between 1st January 2004 and 31 December 2018. We will include individuals with at least 12 months of research standard follow up in CPRD. Therefore, follow up will begin at the latest of 01/01/2004, 65th birthday or 12 months after research standard follow up.

We will follow individuals to the earliest of: incident dementia diagnosis, date of death, transfer out date, the practice's last data collection date or end of study period.

Exclusion
We will exclude individuals with a history of dementia and cognitive impairment. To account for delirium, which is an acute complication of common infections, we will exclude the first 3 months after infection.

K. Selection of comparison group(s) or controls

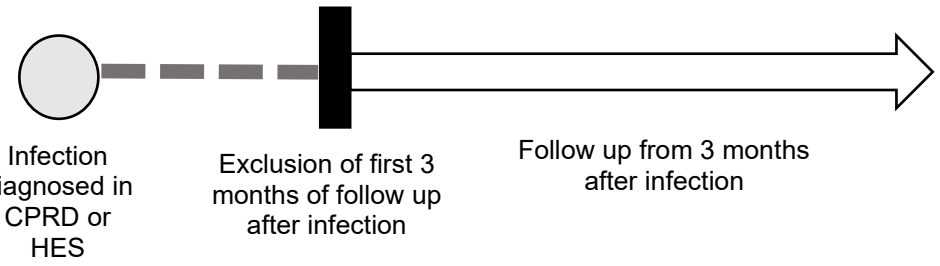
The comparison group will comprise of adults aged 65 and over unexposed to common infections during the study period.

L. Exposures, Outcomes and Covariates

Exposure

Common infections

We will identify Read codes and ICD-10 codes of common infections in both CPRD and HES data.
Common infections can result in short term reversible changes in cognition, delirium, as a result it is possible that individuals diagnosed with dementia shortly after infection could have been experiencing delirium and were misdiagnosed as having dementia. Therefore, the reduce the risk of misclassifying delirium as dementia, we will exclude the first 3 months of follow up after an infection.



Our exposure will reflect an ever diagnosis of infection. This will mean that individuals can move from the unexposed group to exposed but once diagnosed with infection they cannot move to the unexposed group.

Exposure will be defined as one of the following:

1. A clinical code for lower respiratory tract infections
2. A clinical code for sepsis
3. A clinical code for urinary tract infections and a prescription for antibiotics
4. A clinical code for soft tissue infections and a prescription for antibiotics

We will group all common infections (lower respiratory tract, urinary tract, skin and soft tissue infections and sepsis) into one category ‘any infection’ in order to determine the overall association of common infections with incident dementia. Then we will group infections according to subtype of infection. Each subtype of infection will be further subdivided according to the frequency of infections.

Primary Outcome

Dementia

Our primary outcome will be defined as first ever dementia diagnosis. Incident dementia will be identified using Read codes for dementia (any dementia subtype) and ICD-10 codes in HES data. We will exclude those with a prior history of dementia, evidence of cognitive impairment and cases in which dementia occurs before infection.

Secondary Outcome

Evidence of cognitive impairment

Our secondary outcome will be defined as first ever evidence of cognitive impairment. We will identify cognitive impairment using Read codes and ICD-10 codes in CPRD and HES. We will exclude individuals with a prior history of cognitive impairment and dementia.

Covariates

- Age (65-69, 70-74, 75-79,80-84, 85-89, 90+ using data from CPRD)
- Sex (male and female using data from CPRD)
- Ethnicity (White, black, South Asian and other using data from CPRD or HES)
- Calendar year (2004-2008, 2009-2013, 2014-2018)
- Socioeconomic deprivation (using Index of Multiple Deprivation Quintiles)

- BMI, kg/m^2 (underweight, <18.5, normal weight 18.5-25, overweight/obese ≥25 using additional files in CPRD)
- Smoking (Non-smoker, ex-smoker and current smoker using additional files in CPRD)
- Alcohol consumption (non-drinker, current drinker, heavy drinker, light drinker, moderate drinker and ex drinker, using additional files in CPRD)
- Frequency of health service usage (Information relating to health service usage will be obtained from the number of GP consultations and hospitalisations using CPRD and HES)

We will identify read codes relating to the following conditions in clinical, test and therapy files using CPRD and HES:

- Cardiovascular disease: atrial fibrillation, angina, previous myocardial infarction, hypertension, ischemic heart disease, congestive heart failure
- Comorbid conditions: traumatic brain injury, stroke, atherosclerosis, chronic kidney disease, peripheral vascular disease, retinopathy, neuropathy, chronic liver disease, asthma, chronic obstructive pulmonary disease, epilepsy, Parkinson's disease, obstructive sleep apnoea and HIV
- Psychiatric comorbidity: cognitive impairment, post-traumatic stress disorder, major depression, schizophrenia, bipolar disorder and anxiety disorder
- Glycaemic control using HbA1c (<6%, 6-6.5%, 6.5-7%, 8-10% and >10%)

M. Data/ Statistical Analysis

Aim 1
We will describe the age specific incidence rates of dementia in individuals with and without common infections by calculating the number of events and person time at risk of dementia. Age will be stratified into the following groups: 65-69, 70-74, 75-79,80-84, 85-89, 90+.

Aim 2
We acknowledge that the competing risk of mortality is possible when estimating the risk of dementia, particularly in an elderly multimorbid population, and that failure to account for this may lead to biased effect estimates if a competing risk analysis approach is not used. However, when addressing aetiological research questions, Cox regression is an appropriate method of analysis [34, 35]. Therefore, as we aim to investigate the causal relationship between common infections and dementia, we will use Cox regression analysis to estimate the incidence of dementia in those exposed and unexposed to any common infection. Current age will be fitted as the underlying time scale. Our final model will adjust for all the confounders listed previously.
We will then stratify by:

- type of infections (sepsis, lower respiratory tract, urinary tract, skin and soft tissue infections) and this will also include the severity of infections (e.g. hospitalised infections vs non-hospitalised infections or severe sepsis)
- time after infection diagnosis (e.g. 3-12 months, 3-24 months, 3-36 months etc)
- frequency of infections (e.g. 0 1, 2, ≥3).

We will consider infections that occur within 28 days of each other as a single episode of infection.

Aim 3
We will investigate the presence of effect modification by fitting an interaction term of diabetes to our Cox regression model and then carrying out likelihood ratio tests.

Aim 4
We will use Cox regression models to estimate the risk for our secondary outcome of cognitive impairment in those exposed and unexposed to common infections.

Sensitivity analyses

1. We will stratify by dementia subtype in order to explore the incidence of dementia according to subtypes of dementia (Alzheimer's disease and vascular dementia).
2. We will stratify by sex to compare the incidence of dementia in men and women.
3. We will repeat our primary analyses to include individuals in CPRD without HES linked data.
4. We will exclude individuals with read codes for dementia that is causally linked with other diseases such as 'HIV associated dementia' and we will instead perform sensitivity analyses only for dementia that is not specifically caused by a particular disease.
5. To more reliably assess whether infections are associated with evidence of cognitive impairment and we will exclude individuals with recodes for symptoms of cognitive impairment and instead only include those with less ambiguous read codes of cognitive impairment such as 'mild cognitive impairment'.
6. We will repeat our primary analyses defining all types of common infections using clinical codes and an antibiotic prescription.
7. We will repeat our primary analyses stratifying according to the time period before death.

N. Plan for addressing confounding
Our final model will include all potential confounders specified in section K.

O. Plans for addressing missing data
We will describe the pattern of missing data present and choose a suitable method for accounting for missing data accordingly. We expect to find missing data on smoking and ethnicity. However, since these data are less likely to be missing at random, we will most likely use a complete case analysis to carry this out.

P. Patient or user group involvement (if applicable)

This study is funded by the Alzheimer's Society. Through the Alzheimer's society, we have been assigned a group of three lay volunteers who will act as research monitors for the present study. We will seek the research monitors views on the design and conduct of our study as well as the dissemination of our results.

Q. Plans for disseminating and communicating study results, including the presence or absence of any restrictions on the extent and timing of publication

We will disseminate our findings at relevant conferences, events, meetings and we plan to submit our results for publication in a peer reviewed journal. We will work with the Alzheimer's Society to present our findings to members of the public.

Conflict of interest statement: There are no competing interests to declare.

R. Limitations of the study design, data sources, and analytic methods

Misclassification of dementia

There are a number of ways dementia could be misclassified. First, dementia is known to be frequently underdiagnosed in primary care with studies suggesting that over 50% of dementia cases are not detected in primary care [36] [37], although this has been changing with time across this period as recent evidence now suggests that around two thirds of people with dementia have a diagnosis in primary care[38]. Nevertheless, misclassification and underestimation of dementia incidence is possible. However, the positive predictive values of dementia in CPRD are over 80% and we will link CPRD to HES data which will enable us to capture more dementia cases, although this will likely introduce other biases as certain groups such as ethnic minorities and those with milder dementia are less likely to be receive a hospital dementia diagnosis [39]. Therefore, misclassification in HES is also possible although the recording of dementia in HES has been increasing since 2008 and the sensitivity and specificity for each person's complete hospital records has been estimated to be around 78% and 92%, respectively [39]. Second, dementia has a long pre-clinical phase and therefore it is possible that individuals classified as not having dementia in CPRD could already be experiencing cognitive decline or already have dementia. In turn, these individuals may be more susceptible to having infections. To address this, we will present the hazard ratios for incident dementia in different time periods after infection. Additionally, in the period before death, older individuals could be at risk of serious cognitive decline which could also increase their likelihood of a dementia diagnosis. As a result, we will explore the proximity of dementia diagnosis to death. Lastly, common infections are known to be associated with delirium, a serious neuropsychiatric syndrome characterised by acute cognitive dysfunction and inattention. It is therefore possible that delirium may be misclassified as dementia. To reduce this, we will exclude all dementia cases occurring within 3 months after an infection. Additionally, we acknowledge that dementia diagnosed shortly after infection, even after delirium has resolved, is less likely to be causally linked to infection due to the long pre-clinical phase of dementia.

Misclassification of cognitive impairment

Read codes for cognitive impairment have not been validated in CPRD and thus misclassification is possible. Read codes related to cognitive impairment may have been assigned without a diagnostic test and it is possible that individuals who were older and of a lower educational background may have been misclassified as having evidence of cognitive impairment. Additionally, individuals in CPRD are unlikely to have had their cognition tested at multiple time points, as a result, without a comparison of previous cognitive ability, misclassification of cognitive impairment is possible. Furthermore, codes relating to symptoms of cognitive impairment may be inaccurate and may not specifically relate to clinical cognitive impairment. To minimise this, we will perform sensitivity analyses for codes that indicate a diagnosis of cognitive impairment rather than symptoms of cognitive impairment.

Misclassification of common infections

There are a number of ways in which misclassification of infections is possible during this study. Firstly, in primary care settings, infections are often diagnosed without microbiological data to confirm diagnosis. Secondly, people with less serious infections might be less likely to visit the GP which might also lead to an underestimation of people with infections. Lastly, it is possible that people who will be unexposed to infections during the study period were exposed to common infections before the study which might affect their risk for dementia.

Detection bias

People with diabetes are more likely to visit health services compared to those without, thus potentially increasing their chances of a dementia diagnosis. Diabetes is also a known risk factor for dementia and as such it is possible that people with diabetes might be screened more frequently for dementia which might also increase their chances of a dementia diagnosis. Recently, dementia risk has been included in the NHS health checks programme and diabetes has been identified as a risk factor for dementia. This might also increase the likelihood of people with diabetes to receive a dementia diagnosis.

Missing data

Missing data on confounding variables such as ethnicity, smoking and education are likely. We will describe the pattern of our missing data (whether our data is missing completely at random, missing at random or missing not at random) and choose an appropriate method for dealing with the missing data.

S. References

1. Prince M, Guerchet M, Prina M. Policy Brief for Heads of Government: The Global Impact of Dementia 2013–2050. London: Alzheimer's Disease International 2013 [[Internet] http://www.alz.co.uk/research/G8-policy-brief (Last accessed on 25 Jan 2019)
2. Organization. WH. Dementia: a public health priority 2012 [[Internet] http://www.who.int/mental_health/publications/dementia_report_2012/en/ (Last accessed on 6 Feb 2019)
3. Matthews FE, Stephan BC, Robinson L, et al. A two decade dementia incidence comparison from the Cognitive Function and Ageing Studies I and II. *Nat Commun* 2016;7:11398.
4. Schrijvers EM, Verhaaren BF, Koudstaal PJ, et al. Is dementia incidence declining?: Trends in dementia incidence since 1990 in the Rotterdam Study. *Neurology* 2012;78:1456-63.
5. Grasset L, Brayne C, Joly P, et al. Trends in dementia incidence: Evolution over a 10-year period in France. *Alzheimers Dement* 2016;12:272-80.
6. Qiu C, von Strauss E, Backman L, et al. Twenty-year changes in dementia occurrence suggest decreasing incidence in central Stockholm, Sweden. *Neurology* 2013;80:1888-94.
7. van Bussel EF, Richard E, Arts DL, et al. Dementia incidence trend over 1992-2014 in the Netherlands: Analysis of primary care data. *PLoS medicine* 2017;14:e1002235-e.
8. Satizabal CL, Beiser AS, Chouraki V, et al. Incidence of Dementia over Three Decades in the Framingham Heart Study. *N Engl J Med* 2016;374:523-32.
9. Rocca WA, Petersen RC, Knopman DS, et al. Trends in the incidence and prevalence of Alzheimer's disease, dementia, and cognitive impairment in the United States. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2011;7:80-93.
10. Langa KM. Is the risk of Alzheimer's disease and dementia declining? *Alzheimers Res Ther* 2015;7:34.
11. Matthews F, Brayne C, Medical Research Council Cognitive F, et al. The incidence of dementia in England and Wales: findings from the five identical sites of the MRC CFA Study. *PLoS medicine* 2005;2:e193-e.
12. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol* 2011;10:819-28.
13. Norton S, Matthews FE, Barnes DE, et al. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. *Lancet Neurol* 2014;13:788-94.
14. Livingston G, Sommerlad A, Orgeta V, et al. Dementia prevention, intervention, and care. *Lancet* 2017;390:2673-734.
15. Sochocka M, Zwolinska K, Leszek J. The Infectious Etiology of Alzheimer's Disease. *Curr Neuropharmacol* 2017;15:996-1009.
16. Alonso R, Pisa D, Fernández-Fernández AM, et al. Infection of Fungi and Bacteria in Brain Tissue From Elderly Persons and Patients With Alzheimer's Disease. *Frontiers in aging neuroscience* 2018;10:159-.
17. Maheshwari P, Eslick GD. Bacterial infection and Alzheimer's disease: a meta-analysis. *J Alzheimers Dis* 2015;43:957-66.
18. Iwashyna TJ, Ely EW, Smith DM, et al. Long-term cognitive impairment and functional disability among survivors of severe sepsis. *JAMA* 2010;304:1787-94.
19. Amieva H, Jacqmin-Gadda H, Orgogozo JM, et al. The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study. *Brain* 2005;128:1093-101.
20. Chae JH, Miller BJ. Beyond Urinary Tract Infections (UTIs) and Delirium: A Systematic Review of UTIs and Neuropsychiatric Disorders. *J Psychiatr Pract* 2015;21:402-11.
21. Foley NC, Affoo RH, Martin RE. A Systematic Review and Meta-Analysis Examining Pneumonia-Associated Mortality in Dementia. *Dement Geriatr Cogn Disord* 2015;39:52-67.
22. Tate JA, Snitz BE, Alvarez KA, et al. Infection hospitalization increases risk of dementia in the elderly. *Critical Care Medicine* 2014;42:1037-46.
23. Shah FA, Pike F, Alvarez K, et al. Bidirectional Relationship between Cognitive Function and Pneumonia. *American Journal of Respiratory and Critical Care Medicine* 2013;188:586-92.
24. Mawanda F, Wallace RB, McCoy K, et al. Systemic and localized extra-central nervous system bacterial infections and the risk of dementia among US veterans: A retrospective cohort study. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 2016;4:109-17.
25. Dunn N, Mullee M, Perry VH, et al. Association between dementia and infectious disease: evidence from a case-control study. *Alzheimer Dis Assoc Disord* 2005;19:91-4.
26. Casqueiro J, Casqueiro J, Alves C. Infections in patients with diabetes mellitus: A review of pathogenesis. *Indian Journal of Endocrinology and Metabolism* 2012;16:S27-S36.

27. Abu-Ashour W, Twells L, Valcour J, et al. The association between diabetes mellitus and incident infections: a systematic review and meta-analysis of observational studies. *BMJ Open Diabetes Research & Care* 2017;5.
28. Carey IM, Critchley JA, DeWilde S, et al. Risk of Infection in Type 1 and Type 2 Diabetes Compared With the General Population: A Matched Cohort Study. *Diabetes Care* 2018;41:513-21.
29. Hirji I, Guo Z, Andersson SW, et al. Incidence of urinary tract infection among patients with type 2 diabetes in the UK General Practice Research Database (GPRD). *J Diabetes Complications* 2012;26:513-6.
30. Chatterjee S, Peters SA, Woodward M, et al. Type 2 Diabetes as a Risk Factor for Dementia in Women Compared With Men: A Pooled Analysis of 2.3 Million People Comprising More Than 100,000 Cases of Dementia. *Diabetes Care* 2016;39:300-7.
31. Gudala K, Bansal D, Schifano F, et al. Diabetes mellitus and risk of dementia: A meta-analysis of prospective observational studies. *Journal of diabetes investigation* 2013;4:640-50.
32. Cheng G, Huang C, Deng H, et al. Diabetes as a risk factor for dementia and mild cognitive impairment: a meta-analysis of longitudinal studies. *Intern Med J* 2012;42:484-91.
33. Millett ER, Quint JK, De Stavola BL, et al. Improved incidence estimates from linked vs. stand-alone electronic health records. *J Clin Epidemiol* 2016;75:66-9.
34. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol* 2009;170:244-56.
35. Andersen PK, Geskus RB, de Witte T, et al. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol* 2012;41:861-70.
36. Connolly A, Gaehl E, Martin H, et al. Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. *Aging Ment Health* 2011;15:978-84.
37. Iliffe S, Robinson L, Brayne C, et al. Primary care and dementia: 1. diagnosis, screening and disclosure. *Int J Geriatr Psychiatry* 2009;24:895-901.
38. Alzheimer's Research UK Dementia Statistics Hub. Dementia Diagnosis Rate 2018 [cited 2018. [Internet] https://www.dementiastatistics.org/statistics/diagnoses-in-the-uk/; (Last accessed on 24/05/2019)
39. Sommerlad A, Perera G, Singh-Manoux A, et al. Accuracy of general hospital dementia diagnoses in England: Sensitivity, specificity, and predictors of diagnostic accuracy 2008-2016. *Alzheimers Dement* 2018;14:933-43.

List of Appendices

1. **Provisional code lists for outcomes (CPRD)**
a) Dementia
b) Cognitive impairment

2. **Provisional code lists for exposures (CPRD)**
c) Lower respiratory tract infections
d) Sepsis
e) Urinary tract infections and antibiotics
f) Skin and soft tissue infections and antibiotics