New uses for old drugs: Investigating whether antihypertensives can be repurposed for the prevention of dementia

Venexia M Walker

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Health Sciences.

Bristol Medical School

2019

Word count: 40348

Abstract

Background: Drug repurposing applies existing drugs to novel indications to identify potential treatments in a more rapid and cost-effective manner than traditional drug development. Antihypertensive drugs are priority repurposing candidates for dementia prevention, however current evidence for their use is inconclusive. Furthermore, little research has been conducted into the factors that influence prescribing of licensed dementia drugs, which will be an important consideration should a drug repurposing candidate be identified.

Methods: Join point analysis was used to identify factors that have potentially influenced prescribing trends for the licensed dementia drugs in the Clinical Practice Research Datalink (CPRD). Instrumental variable analysis was then conducted in two ways to assess the potential for repurposing antihypertensives for dementia prevention. The first analysis used physicians' prescribing preference as an instrument, also in the CPRD. While the second analysis used genetic variants to proxy the targets of antihypertensive drugs in a two-sample Mendelian Randomization framework.

Results: Prescriptions of dementia drugs have increased since their launch, but the join point analysis suggested that different classes of drugs have been affected by different factors during this time. The instrumental variable analysis using physician's prescribing preference suggested that small differences exist between antihypertensive drug classes in terms of their effect on dementia prevention, but the magnitude of the differences is smaller than previously reported. Finally, Mendelian randomization provided limited evidence that lowering systolic blood pressure, via antihypertensive drug classes, affected Alzheimer's disease risk. This suggests that if specific antihypertensive drug classes affect the risk of Alzheimer's disease, they may not do so via systolic blood pressure.

Conclusions: This thesis has provided new evidence concerning the potential repurposing of antihypertensives for dementia prevention using causal inference methods. In addition, it has examined factors that may have affected dementia drug prescribing, which may also have implications for repurposed drug candidates.

Acknowledgements

Thank you to my supervisors: Neil Davies, Patrick Kehoe, and Richard Martin for the tremendous support and guidance you have provided.

I would also like to acknowledge Liz Coulthard and Jonathan Sterne for their feedback during my annual reviews.

The power calculator for instrumental variable analysis in pharmacoepidemiology, presented in Chapter 3, was the result of a collaboration with Stephen Burgess and Frank Windmeijer. Many thanks to you both.

Thank you also to George Davey Smith, who co-authored the paper 'Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities', which is the basis of Chapter 4.

Tim Jones extracted the data used in this thesis from the Clinical Practice Research Datalink as described in Chapter 5 and co-authored the accompanying protocol paper – thank you for your hard work.

This work was funded by the Perros Trust. Thank you for this opportunity and the interest that you have taken in my work.

Mum, Dad, Alex and Sam have encouraged and supported me throughout this process. Thank you all.

Declaration

I declare that the work in this thesis was carried out in accordance with the requirements of the University of Bristol's Regulations and Code of Practice for Research Degree Programs and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is my own work. Work done in collaboration with, or with the assistance of other, is indicated as such. Any views expressed in this thesis are those of the author.

Signed:

Date:

Contribution statements

Walker VM, Davies NM, Jones T, Kehoe PG, Martin RM. Can commonly prescribed drugs be repurposed for the prevention or treatment of Alzheimer's and other neurodegenerative diseases? Protocol for an observational cohort study in the UK Clinical Practice Research Datalink. BMJ Open. 2016 Dec 1;6(12):e012044.

Contribution statement: As stated in the manuscript, "VMW, NMD, PGK and RMM contributed to planning the analysis. TJ provided the feasibility numbers and refined the cohort definitions. VMW drafted this protocol and all others edited and revised the manuscript. PGK and RMM were responsible for securing the funding."

Walker VM, Davies NM, Windmeijer F, Burgess S, Martin RM. Power calculator for instrumental variable analysis in pharmacoepidemiology. Int J Epidemiol. 2017 Oct 1;46(5):1627–32.

Contribution statement: VMW, NMD and FW derived the power formula. VMW conducted the simulations, developed the associated tools and drafted the manuscript. All authors were involved in the planning of the analysis, as well as the editing and revising of the manuscript.

Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol. 2017 Dec 1;46(6):2078–89.

Contribution statement: VMW drafted the manuscript. All authors were involved in the planning, editing and revising of the manuscript.

Walker VM, Davies NM, Kehoe PG, Martin RM. What is the impact of regulatory guidance and expiry of drug patents on dementia drug prescriptions in England? A trend analysis in the Clinical Practice Research Datalink. Alzheimer's Research & Therapy. 2018 May 29;10:51.

Contribution statement: As stated in the manuscript, "All authors contributed to planning the analysis. VMW conducted the analysis and drafted the manuscript. All other authors edited and revised the manuscript. PGK and RMM were responsible for securing the funding. All authors read and approved the final manuscript."

Walker VM, Davies NM, Martin RM, Kehoe PG. Comparison of antihypertensive drug classes for dementia prevention. bioRxiv. 2019 Jan 12;517482.

Contribution statement: VMW conducted the analysis and drafted the manuscript. All authors were involved in the planning of the analysis, as well as the editing and revising of the manuscript.

Walker VM, Kehoe PG, Martin RM, Davies NM. Repurposing antihypertensive drugs for the prevention of Alzheimer's disease: a Mendelian Randomization study. Int J Epidemiol. 2019 Jul 4; Advance article.

Contribution statement: VMW conducted the analysis and drafted the manuscript. All authors were involved in the planning of the analysis, as well as the editing and revising of the manuscript.

Signed:

Signed:

Signed:

Signed:

Date:

Contents

List of Tables

List of Figures

Glossary

Abbreviation	Term
2SLS	2-stage least squares
AD	Alzheimer’s disease
AChE	Acetylcholinesterase
AMPC	Average monthly percent change
CALIBREX	Candesartan vs lisinopril effects on the brain [trial]
CEDAR	Candesartan's effects on Alzheimer's disease and related biomarkers [trial]
CHD	Coronary heart disease
CI	Confidence interval
CPRD	Clinical practice research datalink
DNA	Deoxyribonucleic acid
EPAD	European Prevention of Alzheimer’s Dementia [consortium]
FOURIER	Further Cardiovascular Outcomes Research With PCSK9 Inhibition in Subjects With Elevated Risk [trial]
GP	General practitioner
GPRD	General practice research database
GWAS	Genome wide association study
HEART	Health evaluation in African Americans using RAS therapy [trial]
HES	Hospital episode statistics
HMGCR	3-hydroxy-3-methylglutaryl-coA reductase
HRT	Horomone replacement therapy
ICD	International classification of diseases
InSIDE	Instrument strength independent of direct effect
ISAC	Independent Scientific Advisory Committee
IV	Instrumental variable
LDL	Low density lipoprotein
MMSE	Mini mental state examination
MPC	Monthly percent change
MR	Mendelian randomization
NHS	National Health Service

NICE	National Institute for Health and Care Excellence
NMDA	N-methyl-d-aspartate
NOS	Not otherwise specified
ONS	Office of National Statistics
OR	Odds ratio
PCSK9	Proprotein convertase subtilisin–kexin type 9
PSD	Potassium-sparing diuretic
PheWAS	Phenome wide association study
QOF	Quality and outcomes framework
RADAR	Reducing pathology in Alzheimer's disease through angiotensin targeting [trial]
RAS	Renin-angiotensin system
RCT	Randomized controlled trial
rrAD	Risk reduction for Alzheimer's disease [trial]
SARTAN-AD	Telmisartan vs perindopril in hypertensive mild-moderate Alzheimer's disease patients [trial]
SNP	Single nucleotide polymorphism
SD	Standard deviation
UK	United Kingdom
USA	United States of America
VAMP Health	Value Added Medical Products Health

Chapter 1. Introduction

1.1. Summary

This chapter presents the problem that the research conducted in this thesis seeks to address and an outline of the main output of this thesis: new evidence concerning repurposing antihypertensives for dementia prevention using genetic and non-genetic instrumental variable analyses. This chapter then introduces three issues that arose and have been addressed during the conduct of this research. They are: how to calculate power for instrumental variable analyses in pharmacoepidemiology; how to use Mendelian randomization to predict drug repurposing opportunities; and what factors have effected prescribing of existing dementia drugs. This chapter concludes with the aims and objectives of this thesis, a summary of its organization and a list of the associated outputs.

1.2. Statement of the problem

There is a substantial unmet clinical need for treatments for dementia where significant benefits to patients, society and the public purse can be gained. Despite this, some drug companies have recently withdrawn from this therapy area due to failed and costly efforts to find new treatments. (1,2) Drug repurposing, the identification of properties in existing or abandoned compounds for other clinical conditions, offers significant advantages over traditional drug discovery approaches. This includes immediate access to human safety data from the original clinical development work, which can accelerate testing in clinical trials, saving both time and money. (3–5)

Many antihypertensive medications have been proposed as drug repurposing candidates for the prevention of dementia. In part, because of research to better understand several reports of observed associations between midlife hypertension and later-life risk of Alzheimer's disease and vascular dementia. (5–8) There is also increasing recognition that one of the earliest pathological events in the development of Alzheimer's disease is vascular dysregulation. (9) As well as suggestions that some antihypertensives, specifically those that block angiotensin receptor and calcium channel signalling, may have other neurological benefits. (9–11)

Several observational studies have investigated repurposing antihypertensives for dementia prevention. (12–19) However, these studies have typically used case-control designs with logistic regression and cohort designs with survival analysis, both of which are observational study designs that may be subject to unmeasured or residual confounding and reverse causation. Specific concerns of observational studies are confounding by indication, where the reasons that a patient receives a treatment relate to the reasons that the patient is at an increased risk of the outcome; and healthy adherer bias, where patients initiating or adhering to a drug for prevention of a condition are more likely to be healthy. There is also potential for reverse causation due to preclinical or early stages of the disease, which could lead to more frequent contacts with general practitioners (GPs) and lead to raised blood pressure being more likely to be detected. This, in turn, could lead to the prescription of an antihypertensive drug in advance of dementia being formally diagnosed. Consequently, current evidence concerning repurposing antihypertensives for dementia prevention is considered inconclusive.

1.3. New evidence concerning repurposing antihypertensives for dementia prevention using instrumental variable analysis

Instrumental variable analysis, which estimates the causal effect of an exposure on an outcome by using a third variable (the instrument), can be robust to confounding and reverse causation if certain assumptions are met. That is:

IV1. The instrument must associate with the exposure

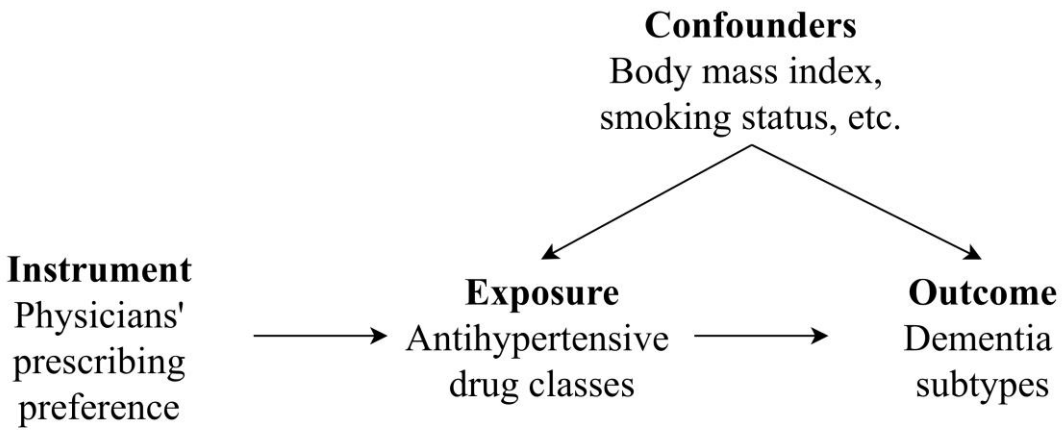
IV2. The instrument must only affect the outcome through the exposure

IV3. The instrument and the outcome must have no common causes

These assumptions can be represented on a directed acyclic graph, as shown in Figure 1.1. Instrumental variable analysis, with other sources of evidence, can be used in a triangulation framework to obtain a reliable answer concerning the potential repurposing of antihypertensives for dementia prevention. (20) This thesis presents two forms of instrumental variable analysis to provide new evidence concerning this hypothesis. The first uses physicians' prescribing preference as a non-genetic instrument in electronic health record data obtained from the United Kingdom (UK) Clinical Practice Research Datalink (CPRD). The second uses single nucleotide polymorphisms (SNPs) i.e. differences in the deoxyribonucleic acid (DNA) nucleotides between individuals, which have been selected to mimic the biological function of the protein targets of

antihypertensive drug classes, as a genetic instrument in an approach more commonly known as Mendelian randomization. The results of both these instrumental variable analyses have been made available via the following references. (21,22)

Figure 1.1: A directed acyclic graph illustrating the basic instrumental variable analysis model.



Directed acyclic graphs are a visual representation of a model, which represent variables by 'nodes' and the relationships between them by 'directed edges'. (68) The graphs are defined as acyclic because edges cannot form 'cycles', whereby the edges all act in the same direction. This prevents a variable from being both a cause and a consequence of itself. Directed acyclic graphs are a common tool in epidemiology and, more specifically, causal inference when edges are given a causal interpretation, as they allow researchers to depict their model and its assumptions using a common graphing scheme. (69)

1.4. Calculating power for instrumental variable analysis in pharmacoepidemiology

Instrumental variable analysis is an increasingly popular method in the field of pharmacoepidemiology. (23–28) However, the power calculators that were available for studies using instrumental variable analysis at the start of my PhD – such as Mendelian randomisation power calculators – did not allow for the structure of research questions using non-genetic instruments (for example, physicians' prescribing preference) in this field. (29,30) This is because analysis using non-genetic instruments in pharmaco-epidemiology will typically have stronger instruments and so can detect smaller causal effects. Consequently, there was a need for dedicated power calculators for these type of research questions in pharmacoepidemiology. In this thesis, I investigate how to conduct

power calculations for pharmacoepidemiological studies, which use a single binary instrument to analyse the causal effect of a binary exposure on a continuous outcome. I also provide an online calculator, as well as packages in both R and Stata, for the implementation of the formula by others (https://github.com/venexia/PharmIV). This work has been published in the International Journal of Epidemiology. (31)

1.5. Using Mendelian randomization to predict drug repurposing opportunities

Identification of drug repurposing opportunities can maximize the benefit of a drug. However, as highlighted before, the more traditional observational research methods used to investigate these opportunities are subject to several biases. These include confounding by indication, reverse causality, and missing data. In this thesis, I propose Mendelian randomization as a novel approach that can be used for the prediction of drug repurposing opportunities. Mendelian randomization addresses some of the limitations associated with the existing methods in this field. Furthermore, it can be applied either pre- or post-approval of the drug and could therefore prevent the potentially harmful exposure of patients in clinical trials and beyond. This thesis includes discussion of examples from the literature that have used Mendelian randomization to predict drug repurposing opportunities and covers the strengths and limitations associated with using this method for this purpose. There was relatively little discussion focussed on using Mendelian randomization for drug repurposing when I commenced my PhD and I have since published on this topic. (32)

1.6. Factors effecting existing dementia drug prescribing

Drugs for dementia have been available in England from 1997. However, since their launch, there have been several changes to national guidelines and initiatives that may have influenced prescribing. These include changes in National Institute for Health and Care Excellence (NICE) guidance; several government dementia strategies; the addition of dementia to the Quality and Outcomes Framework (QOF); and the expiry of drug patents. Despite this, little research had been conducted prior to my PhD into the effect of these events on prescribing. (33,34) In this thesis, I investigate prescribing trends in England since the launch of these drugs up to 1st January 2016 using data from the CPRD to address this gap in the literature. The key motivation for this analysis was to

identify factors that have affected prescriptions of existing treatments that may also influence repurposed drug candidates in the future. However, the results from this analysis could also be used to identify breaks in the prescription of these drugs that could be exploited as natural experiments for progression studies. For example, if these drugs were not prescribed when the NICE guidelines stopped recommending their use between 2006 and 2011, the progression of people diagnosed during this time who did not access the drugs could be compared with the progression of people diagnosed before and after this time who did have access to the drugs. The results presented in this chapter have been published in Alzheimer's Research & Therapy. (35)

1.7. Aims and objectives

The objective of this thesis was to use instrumental variable analysis methods, in existing data sources, to triangulate evidence for repurposing antihypertensive drugs for the prevention of dementia.

The specific aims were as follows:

1. Develop a power calculator for non-genetic instrumental variable analysis studies in the context of pharmacoepidemiology.
2. Describe the use of genetic instrumental variable analysis, namely Mendelian randomization, for predicting drug repurposing opportunities.
3. Examine the impact of regulatory guidance and patent expiry on dementia drug prescribing.
4. Investigate whether antihypertensive drugs have a causal effect on incident dementia using instrumental variable analysis with electronic health record data.
5. Investigate whether antihypertensive drugs have a causal effect on incident dementia using instrumental variable analysis with genetic data.

1.8. Organization of this thesis

Chapter 2 provides the background necessary for the rest of this thesis through the introduction of dementia, the concept of drug repurposing, and discussion of the existing evidence regarding antihypertensive drugs for dementia prevention. It also covers the

strengths and limitations of observational pharmacoepidemiology and explains why it might be preferable over 'gold standard' randomized controlled trials for some hypotheses. Chapter 3 describes instrumental variable analysis, the method utilized throughout this thesis, and documents the development of a power calculator for this method in the context of pharmacoepidemiology (Aim 1). Chapter 4 introduces the idea of using Mendelian randomization, a form of instrumental variable analysis that uses genetic variants as instruments, for drug repurposing and covers the strengths and limitations of this approach (Aim 2). Chapter 5 describes the CPRD and my use of this data source. Chapter 6 covers the current treatments available for dementia and factors affecting their prescription in England based on data from the CPRD (Aim 3). Chapter 7 presents an assessment of the effects of antihypertensive drugs on dementia prevention using instrumental variable analysis with data from the CPRD (Aim 4); while Chapter 8 presents an assessment of the effects of antihypertensive drugs on dementia prevention using instrumental variable analysis with genetic data (Aim 5). The thesis concludes with a discussion in Chapter 9 that brings together all the elements of this thesis and discusses their implications.

1.9. Outputs from this thesis

1.9.1. Contributions to scientific literature

Contributions to scientific literature arising from this thesis are detailed below and provided in Appendix A.

The protocol for the observational work using CPRD data to investigate drug repurposing opportunities is published in the BMJ Open. (36) Its contents are referenced in several places throughout this thesis, particularly Chapter 5:

Walker VM, Davies NM, Jones T, Kehoe PG, Martin RM. Can commonly prescribed drugs be repurposed for the prevention or treatment of Alzheimer's and other neurodegenerative diseases? Protocol for an observational cohort study in the UK Clinical Practice Research Datalink. BMJ Open. 2016 Dec 1;6(12):e012044.

The power calculator for instrumental variable analysis in pharmacoepidemiology, described in Chapter 3, is published in the International Journal of Epidemiology (31):

Walker VM, Davies NM, Windmeijer F, Burgess S, Martin RM. Power calculator for instrumental variable analysis in pharmacoepidemiology. Int J Epidemiol. 2017 Oct 1;46(5):1627–32.

Also published in the International Journal of Epidemiology is an article discussing Mendelian randomization as a novel approach for the prediction of adverse drug events and drug repurposing opportunities (32), which formed the basis of Chapter 4:

Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol. 2017 Dec 1;46(6):2078–89.

The trend analysis examining prescribing practice for drugs for dementia in the CPRD, presented in Chapter 6, is available from Alzheimer's Research & Therapy (35):

Walker VM, Davies NM, Kehoe PG, Martin RM. What is the impact of regulatory guidance and expiry of drug patents on dementia drug prescriptions in England? A trend analysis in the Clinical Practice Research Datalink. Alzheimer's Research & Therapy. 2018 May 29;10:51.

The assessment of antihypertensives for dementia prevention using electronic health record data, reported in Chapter 7, is currently under peer review and available from bioRxiv (21):

Walker VM, Davies NM, Martin RM, Kehoe PG. Comparison of antihypertensive drug classes for dementia prevention. bioRxiv. 2019 Jan 12;517482.

The assessment of antihypertensives for dementia prevention using genetic data, reported in Chapter 8, has been made available as an advance article from the International Journal of Epidemiology (22):

Walker VM, Kehoe PG, Martin RM, Davies NM. Repurposing antihypertensive drugs for the prevention of Alzheimer's disease: a Mendelian Randomization study. Int J Epidemiol. 2019 Jul 4; Advance article.

1.9.2. Contributions to scientific meetings

I presented the paper "Power calculator for instrumental variable analysis in pharmacoepidemiology", described in Chapter 3, at the UK Administrative Data Research Network Annual Research Conference 2017 in Edinburgh, UK.

I presented the paper "Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities", which formed the basis of Chapter 4, at the University of Bristol Population Health Symposium 2016 in Bristol, UK and at the International Society for Pharmacoepidemiology mid-year meeting 2018 in Toronto, Canada.

I presented posters, based on the paper "What is the impact of regulatory guidance and expiry of drug patents on dementia drug prescriptions in England?" detailed in Chapter 6, at the Alzheimer's Research UK Research Conference 2016 in Aberdeen, UK and at the University of Bristol brain research showcase and networking day 2018 in Bristol, UK.

I presented "Can treatments for hypertension be repurposed for the treatment of dementia?" at the Society of Epidemiologic Research annual conference 2018 in Baltimore, United States of America (USA); at the European Congress of Epidemiology 2018 in Lyon, France and the International Society for Pharmacoepidemiology annual meeting 2018 in Prague, Czech Republic. These presentations combined results from Chapters 7 and 8.

Chapter 2. Background: Dementia, antihypertensive drugs and observational pharmacoepidemiology

2.1. Introduction

This chapter has three parts, the first of which concerns dementia. Specifically, I describe the public health problem dementia poses, the epidemiology of the condition and the currently available treatments. The second concerns drug repurposing. This part starts with an introduction to the concept of drug repurposing; followed by a description of the antihypertensive drugs that are of interest for this thesis; and the existing evidence for them as drug repurposing candidates for dementia prevention. The final part concerns observational pharmacoepidemiology. This covers the strengths and limitations compared with randomized clinical trials, which are considered the gold standard, and the motivation for using instrumental variable designs to analyse observational data in this thesis. This leads on to Chapters 3 and 4, which formally introduce instrumental variable analysis using electronic health record data and genetic data respectively.

2.2. Public health problem

There are estimated to be 50 million people living with dementia worldwide. (37) This includes 850,000 people in the UK. (38) As of 2015, dementia was the leading cause of death in the UK and "dementia is the only condition in the top 10 causes of death without a treatment to prevent, cure or slow its progression". (39,40) This is despite investment in 1120 unique drug targets between 1995 and 2014. (3,5,41) Consequently, as stated in Section 1.2, there is a substantial unmet clinical need for new treatments for dementia where benefits to patients, society and the public purse can be gained.

2.3. Epidemiology of dementia

Dementia is a progressive condition, which describes several symptoms relating to memory loss and difficulties with thought or speech. The condition results from damage to the brain by other diseases and so has several different forms depending on the cause of damage. The most common form of dementia is Alzheimer's disease, which accounts

for approximately 62% of all cases, but other common forms include vascular dementia (~17%), mixed dementia (~10%) and Lewy body dementia (~4%). (42)

The pathology of Alzheimer's disease is characterised by structural changes to the brain, specifically the formation of amyloid plaques and neurofibrillary tangles. The amyloid cascade hypothesis suggests that these structures form due to over-production of neurotoxic forms of amyloid beta. The excess amyloid beta causes damages to nerve cells and forms plaques. These changes in turn cause modifications to the protein tau, which is associated with the building of microtubules – an important part of the structural integrity of neurons. Consequently, tau aggregates to form neurofibrillary tangles that interfere with cell function and health. The mechanisms and pathways linking amyloid beta and tau still remain unclear, but are the focus of much research intended to find effective disease modifying treatments. (43,44)

Several cardiovascular risk factors in midlife have been linked with the development of Alzheimer's disease in later life, though this relationship is not yet fully understood. (45–47) As the causes of Alzheimer's disease remain unknown, it is thought that risk factors may help to provide a greater insight into the biological mechanisms of the disease. Hypertension, hypercholesterolaemia, and type 2 diabetes are of interest as risk factors due to the large number of people diagnosed with these conditions and the availability of commonly prescribed drugs that could be potentially repurposed (Section 2.5). This thesis will focus on hypertension.

2.4. Current dementia treatments

There are currently four licensed treatments that provide symptomatic relief for patients with Alzheimer's disease in the UK – three acetylcholinesterase inhibitors (donepezil, rivastigmine, galantamine) and one N-methyl-D-aspartate (NMDA) receptor antagonist (memantine). These drugs are collectively referred to as drugs for dementia in the British National Formulary but are licensed for use in Alzheimer's disease only. (48) They are currently taken by approximately 54%, 72% and 13% of people with mild, moderate, and severe dementia respectively. (49) This was estimated to cost £28 million for the year 2017 in England alone. (50) There are no drugs currently licensed for the symptomatic relief of other forms of dementia, though there are unlicensed indications relating to

dementia with Lewy bodies or mixed dementias that include Alzheimer's disease listed in the British National Formulary. (49) At present, there are still no disease-modifying or preventative drugs for any form of dementia. Plus, given the large number of failed attempts in this therapy area, the success rate for Alzheimer's disease drug development is estimated to be just 0.5%. (3,5,41) This has led to some drug companies withdrawing from drug development for dementia. (1) As well as, the formation of ambitious large-scale commercial and non-commercial partnerships, such as the European Prevention of Alzheimer's Dementia (EPAD) Consortium, to try and overcome the cost burden to any single organisation. (51)

2.5. Drug repurposing

Given the lack of current treatments for dementia, and the significant cost of traditional drug development endeavours, drug repurposing is a promising alternative. The medicines and healthcare product regulatory agency suggest that traditional drug development efforts can take up to 15 years and more than a billion pounds to develop a drug from a chemical compound to the point where it can be sold on the pharmacy shelf. (52) Drug repurposing, the testing and use of existing drugs for new indications, is considered a time- and cost-effective alternative to these traditional methods. The key advantage of this technique is that the drugs considered have already been subject to the initial testing required for regulatory approval and, in some cases, already exist in generic form and so are highly economical. Consequently, drug repurposing is growing in popularity and many areas of medicine have already benefited from its application; including cancer, smoking cessation, obesity and Parkinson's disease. (4,5)

Drug repurposing opportunities can be discovered throughout the drug development process. However, prior to the approval of a novel drug, its risk-benefit profile cannot be fully known. This is because pre-approval clinical trials are principally for demonstrating the drug's efficacy for its intended indication. This limits the trial's ability to assess safety and identify novel indications in a number of ways. (53) Firstly, the comparatively small number of patients exposed to a drug during a pre-approval clinical trial means that only very common or very large drug effects can be detected. Secondly, the length of time that patients are exposed to the drug in this setting is relatively short. Thirdly, the recorded data may not include the necessary information to identify previously unknown drug

effects or those that are unrelated to the drug's indication. Finally, the participants of a study may not represent the broad range of patients seen in clinical practice or may be limited in some way – for example, phase I trials have previously been restricted to male participants only. (54) As a result of these limitations, continued assessment of drugs post-approval, as is conducted in this thesis, is necessary in order to fully develop their profile and identify potential drug repurposing opportunities.

2.6. Antihypertensive drugs

This thesis will focus on whether antihypertensive drugs can be repurposed for the prevention of dementia. Hypertension is a mostly symptomless, chronic illness characterised by blood pressure in excess of 140/90 mm/Hg. A 2014 health survey for England found the prevalence of hypertension to be 32% among men and 27% among women surveyed. (55) Most of these cases are examples of 'essential hypertension' – this means the cause of excess pressure is unknown. Though the cause of essential hypertension is unknown, the condition is a known risk factor for many other diseases, including Alzheimer's disease and vascular dementia. (56,57) This is thought to be due to the strain it puts on the heart and vascular system, which can in turn cause or relate to the development of cerebrovascular disease and, in a proportion of cases, different forms of stroke. Treatments for hypertension are among the most commonly prescribed in primary care and are used by 83% of men and 89% of women who are diagnosed with the condition, according to the 2014 health survey for England. (55,58) There are several drug classes available for the treatment of the condition, including beta-adrenoceptor blockers, calcium channel blockers, a range of diuretics, and drugs affecting the renin-angiotensin system. These drug classes have variable effects on blood pressure but are reported, on average, to reduce systolic blood pressure by 9mmHg. (59)

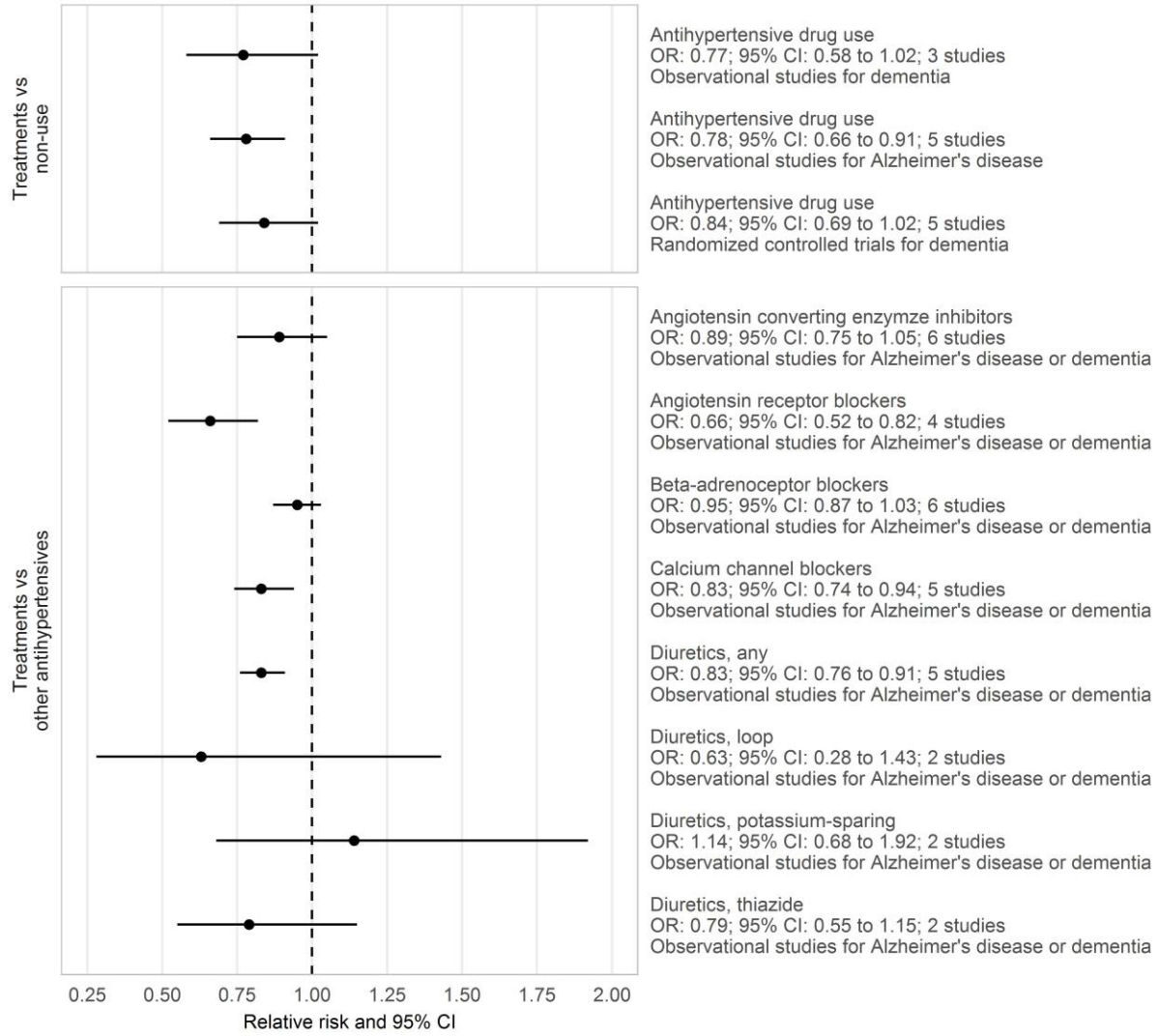
Many antihypertensive drugs have been proposed as drug repurposing candidates for the prevention of dementia. Most notably, they were one of the key treatment groups highlighted in a consensus study of experts conducted by Corbett et al in 2012. (5) As discussed in Section 1.2, there are several reasons for this. These include the observed associations between midlife hypertension and later-life risk of Alzheimer's disease and vascular dementia (5–8); the increased recognition of the role of vascular dysregulation in Alzheimer's disease (9); and suggestions of other neurological benefits, independent of

blood-pressure lowering and related to other pathological processes in Alzheimer's disease, for certain drug classes. (9–11)

2.7. Current evidence concerning repurposing antihypertensives

A systematic review and meta-analysis conducted in mid-2018 by Larsson et al summarized the evidence concerning repurposing antihypertensives for dementia prevention available at that time. (12) The results from this meta-analysis are presented in Figure 2.1. Larsson et al identified five randomized controlled trials and 13 prospective studies that had compared antihypertensives against non-use for dementia. They also identified a further seven observational studies that had compared antihypertensive drug classes against each other, but no trials. (12–19) Four out of the five trials that compared antihypertensives against non-use had point estimates that suggested a protective effect, however three of these four trials failed to exclude the null. This resulted in the meta-analysis finding an overall relative risk of 0.84 (95% confidence interval (CI): 0.69 to 1.02). The results from observational studies were similar with a relative risk of 0.77 (95% CI: 0.58 to 1.02) for dementia based on three comparable studies and 0.78 (95% CI: 0.66 to 0.91) for Alzheimer's disease based on five comparable studies. The evidence for comparing antihypertensives was mixed with one study (of three) suggesting angiotensin-converting enzyme inhibitors were protective (13–15); three studies (of four) suggesting angiotensin-II receptor blockers were protective (15–18); and one study (of one) suggesting calcium channel blockers were protective (19) when compared against other antihypertensive drug classes. Larsson et al concluded their review by stating "available evidence from RCTs and prospective studies indicates that antihypertensive therapy might have a role in preventing dementia and AD". However, there are several issues with the evidence collected to date. In particular, the fact that most of the trials included in the meta-analysis were from populations with high cardiovascular morbidity and were designed around cardiovascular related primary outcomes. This could mean that the proportion of dementia cases that derived from vascular mechanisms will be disproportionately high compared with other populations. (60,61)

Figure 2.1: Summary of results concerning the use of antihypertensives to prevent dementia and Alzheimer's disease from the Larsson et al (2018) meta-analysis.



Note that the number of studies in the meta-analysis may differ from the number of studies reported in the systematic review. Reproduced from Larsson, S. C., & Markus, H. S. (2018). Does Treating Vascular Risk Factors Prevent Dementia and Alzheimer's Disease? A Systematic Review and Meta-Analysis. Journal of Alzheimer's Disease, 64(2), 657–668. https://doi.org/10.3233/JAD-180288

Since the publication of this systematic review and meta-analysis, research has continued into the possible repurposing of antihypertensives for dementia prevention and two key new studies have been published. The first was a study by Barthold et al that compared Alzheimer's disease incidence between users of renin-angiotensin system acting drug classes and non-renin-angiotensin system acting drug classes across sex, race, and ethnic groups in the USA. (62) This study found that angiotensin-II receptor blockers may reduce the risk of Alzheimer's disease in certain groups, namely white and black women and white men. This study is important because evidence to date for the use of these drugs to prevent dementia in non-white populations has been limited; renin-angiotensin system acting drugs are reported to function differently in non-Caucasians; and rates of dementia are higher among African Americans. (62) The second key study was the SPRINT-MIND trial, which assessed the effect of intensive vs standard blood pressure control using a range of antihypertensive medications on probable dementia, mild cognitive impairment and a composite outcome combining probable dementia and mild cognitive impairment, have also been released. (63) The trial found evidence to suggest that intensive blood pressure control was beneficial for the mild cognitive impairment and composite outcomes. Meanwhile, the estimate for the primary cognitive outcome of probable dementia included the null within its CI but may have been underpowered due to the early termination of the trial. There are also a number of trials concerning repurposing antihypertensives for dementia prevention currently ongoing, which are summarized in Table 2.1.

Given the lack of conclusive trial evidence published to date, and the limitations of the conventional observational study designs (Section 2.8.1) used to date, further research into the potential repurposing of antihypertensive drugs for the prevention of dementia is warranted.

2.8. Observational pharmacoepidemiology

As described in the previous section, several observational studies have investigated repurposing antihypertensives for dementia prevention to date. However, these studies have used case-control designs with logistic regression and cohort designs with survival analysis, which may be subject to certain biases. Below I describe several key limitations of observational pharmacoepidemiology, which may have affected previous studies. This

is followed by several strengths, which will help motivate my choice to use observational data and causal inference to provide new evidence concerning the potential repurposing of antihypertensives for dementia prevention. The strengths and limitations are summarized in Table 2.2.

Table 2.1: Currently ongoing trials concerning repurposing antihypertensives for dementia prevention.

Trial Name	Trial ID	Treatment comparison
Candesartan vs lisinopril effects on the brain (CALIBREX)	NCT01984164	Lisinopril vs candesartan
Candesartan's effects on Alzheimer's disease and related biomarkers (CEDAR)	NCT02646982	Candesartan vs placebo
Health evaluation in African Americans using RAS therapy (HEART)	NCT02471833	Telmisartan vs placebo
Reducing pathology in Alzheimer's disease through angiotensin targeting (RADAR)	ISRCTN93682878	Losartan vs placebo
Risk reduction for Alzheimer's disease (rrAD)	NCT02913664	Losartan vs amlodipine
Telmisartan vs perindopril in hypertensive mild-moderate Alzheimer's disease patients (SARTAN-AD)	NCT02085265	Perindopril vs telmisartan

Reproduced from Kehoe, P. G. (2018). The coming of age of the angiotensin hypothesis in Alzheimer's disease: Progress toward disease prevention and treatment? Journal of Alzheimer's Disease, 62(3), 1443–1466. https://doi.org/10.3233/JAD-171119

Table 2.2: Strengths and limitations of observational pharmacoepidemiology.

Strengths	• Allows measurement of effectiveness • Long follow up • Unexpected effects can be studied • Large sample size • Heterogeneous treatment effects can be detected • Reduced cost compared to randomized controlled trials
Limitations	• Ascertainment bias • Collider bias • Immortal time bias • Confounding by indication • Time dependent confounding

2.8.1. Limitations of observational pharmacoepidemiology

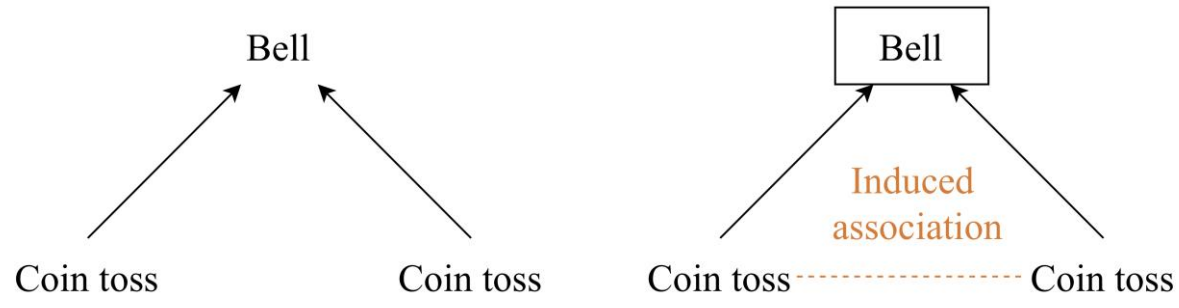
2.8.1.1. Ascertainment bias

Ascertainment bias results from distortion in the number of recorded events. (64) For example: patients with chronic conditions, such as diabetes, are more likely to see a GP frequently. This in turn increases their opportunity of receiving other diagnoses, such as dementia. In this situation, controlling for consultation rate can help to minimize this form of bias.

2.8.1.2. Collider bias

A collider is a variable that is a common effect of two other variables. If the common effect is conditioned on in the analysis, an association is induced between the two variables, leading to bias within the results. (65) This is illustrated in Figure 2.2. An example of collider bias in a pharmacoepidemiological study would be conditioning on an event that happened as a result of a patient's prescription, as this could obscure the effect of the prescribed drug. This is because the measured outcome results from both the prescribed drug and the event that occurred due to the prescribed drug. To prevent this form of bias from affecting results, measures such as only using data inputted prior to the index date to define covariates can be implemented. (66)

Figure 2.2: Illustration of collider bias.

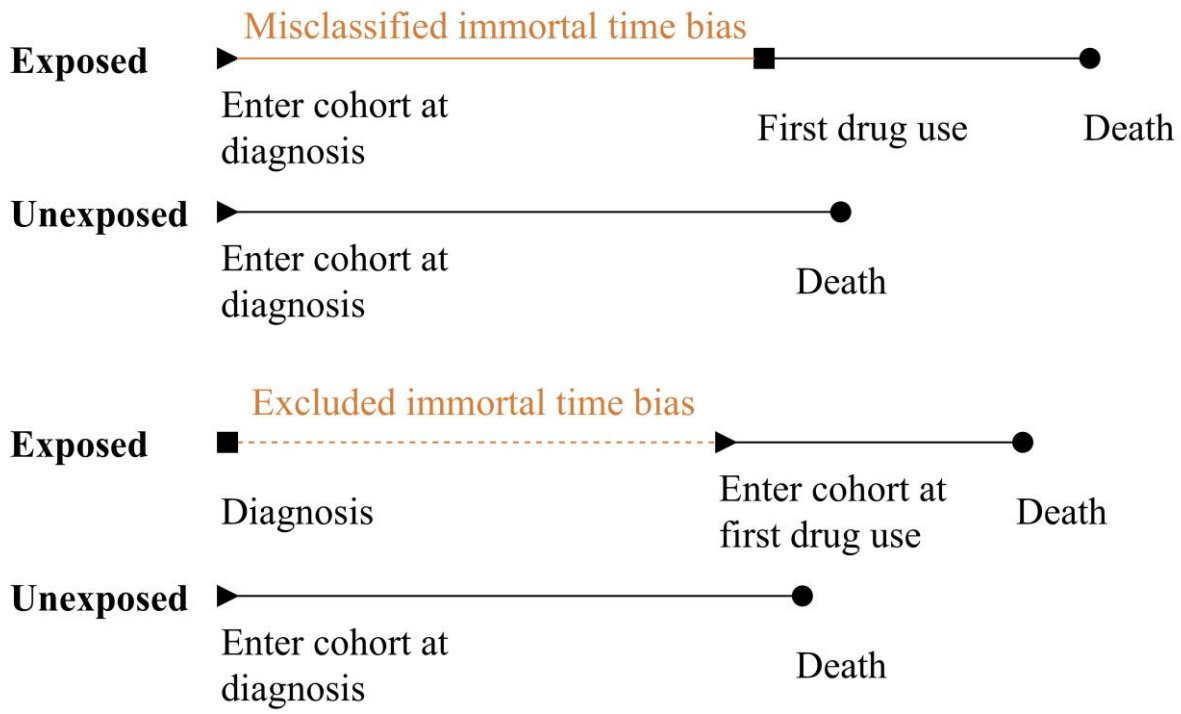


Two coins are tossed independently. If either coin is a head, a bell rings. If we know that one of the coin tosses was a tails and we heard the bell, then we can infer that the other coin toss must have been a head. The bell is therefore a collider and an association is induced between the two independent coin tosses when it is conditioned on. Reproduced from Gage, S. H., Davey Smith, G., Ware, J. J., Flint, J., & Munafò, M. R. (2016). G = E: What GWAS Can Tell Us about the Environment. PLOS Genetics, 12(2), e1005765.

2.8.1.3. Immortal time bias

Immortal time is a period of follow up in which an event cannot occur. This causes bias by altering the ratio of 'exposed' and 'unexposed' time considered in an analysis. Immortal time bias has two forms: excluded immortal time bias or misclassified immortal time bias, which are illustrated in Figure 2.3. (67) Excluded immortal time bias occurs when follow up starts on an inconsistent index date. For example, following a patient from exposure to the treatment of interest ensures they have survived long enough to receive that treatment i.e. they are immortal in the period prior to follow up. To resolve this, follow up should start on a consistent index date, such as diagnosis of the treatment indication, and any time prior to exposure is considered 'unexposed'. Misclassified immortal time bias occurs when periods of follow up are incorrectly categorised as 'exposed' or 'unexposed'. For example, assigning a patient to the 'exposed' group from cohort entry instead of first exposure to the treatment of interest will result in an immortal period between these two dates. To avoid this type of bias, follow up of patients who receive the treatment of interest should include both 'exposed' and 'unexposed' periods according to date of treatment. (67–69)

Figure 2.3: Illustration of misclassified and excluded immortal time bias.

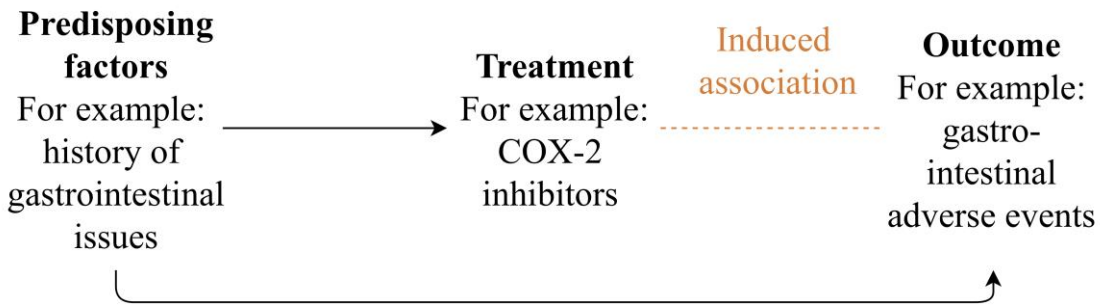


Reproduced from Suissa, S. (2007). Immortal time bias in observational studies of drug effects. Pharmacoepidemiology and Drug Safety, 16(3), 241–249. https://doi.org/10.1002/pds.1357

2.8.1.4. Confounding by indication

Confounding by indication occurs when the factors predisposing a patient to receive treatment are the same factors that predispose them to have an outcome. This induces an artificial association between the exposure and the outcome, as illustrated in Figure 2.4. For example, COX-2 inhibitors are a class of non-steroidal anti-inflammatory drugs that are marketed as being less likely to cause gastrointestinal adverse effects and so are more likely to be prescribed to people at increased risk of such effects. (70,71) Confounding by indication can be minimized by ensuring the exchangeability of patients between the exposed and unexposed groups and using methods such as instrumental variable analysis.

Figure 2.4: Illustration of confounding by indication.



2.8.1.5. Time dependent confounding

Time dependent confounders are confounders that are both caused by the exposure and effected by the exposure. For example, consider comparing two drugs for hypercholesterolaemia with the goal of reducing coronary heart disease risk. Low density lipoprotein (LDL) cholesterol levels will naturally vary over time within an individual. LDL cholesterol will also be used to determine whether treatment is required, whether the treatment has worked, and whether future treatment is required. Consequently, values of LDL cholesterol measured after exposure to the drugs will be causally related to the exposure, which occurred before, and the outcome, which will occur afterwards. This can be difficult to overcome using conventional methods, however methods such as marginal structural models have been developed to overcome this type of confounding. (72)

2.8.2. Strengths of observational pharmacoepidemiology

2.8.2.1. Allows measurement of effectiveness

Efficacy refers to how well a treatment works in the context of a randomized controlled trial, while effectiveness refers to how well a treatment works in the actual population. Randomized controlled trials measure efficacy, as the participants must meet strict criteria in order to be included. Observational studies can be used to estimate effectiveness. The advantage of measuring effectiveness over efficacy is that it gives a measure of how well the treatment would work in 'reality', rather than in an ideal or more highly controlled clinical trial setting. In particular, it allows the study of treatment effects among the general population rather than a randomized controlled trial population, which may not be representative of the general population. This is an important factor when considering the value of a treatment and the guidelines for its prescription.

2.8.2.2. Long follow up

Observational studies can have much longer follow-up than other study designs - this is especially true of observational studies using routinely collected data, such as the CPRD (Chapter 5). The cost of obtaining observational data compared to the cost of obtaining the information in a randomized controlled trial, which will have strict data collection policies and may require specially trained staff, means that the funds for such research go much further. Furthermore, the burden on participants is less, as they are not required to follow the strict treatment plans used in randomized controlled trials but instead operate as they would normally.

2.8.2.3. Unexpected effects can be studied

Observational studies can have the scope to detect unexpected effects as additional data is collected. This is not necessarily the case for randomized controlled trials where only data relevant to the outcome of interest and the covariates is collected. Unexpected effects can be both favourable – such as a drug that improves more than just the

indication it was intended for – or unfavourable – such as a drug with a rare adverse side effect – however both types of effect are important in the evaluation of a drugs' safety.

2.8.2.4. Large sample size

The relatively low cost of data collection in observational studies compared with randomized controlled trials means large sample sizes are achievable. Again, this is especially true of routinely collected data, such as the CPRD (Chapter 5). Large sample sizes have many benefits including greater statistical power to detect treatment effects. This is important as even small treatment effects can have a large clinical impact for patients. A large sample size also allows for the identification of rare outcomes, which can be missed in small samples due to their infrequency. Although rare, these outcomes still need to be considered for treatment safety – especially for commonly prescribed treatments.

2.8.2.5. Heterogeneous treatment effects can be detected

Treatments for which the effect varies according to who receives it are heterogeneous. Observational studies allow the identification of subgroups of a population that may benefit more, or less, from a treatment. This is because they tend to collect data on a broader range of subjects where groups of patients are distinguishable from each other. Treatments with heterogeneous effects can appear to have no effect if the harmful effect caused to one group of patients balances the beneficial effect to another group. This leads to missed beneficial effects that could be improving the lives of patients.

2.8.2.6. Reduced cost compared to randomized controlled trials

Data collection for observational studies can cost much less than the comparable randomized controlled trials. The three main constraints on cost are the number of patients i.e. the sample size; the amount of information collected about a patient i.e. the covariates; and the amount of time a patient is followed i.e. the follow-up. For a given cost, an observational study could provide more information on one or more of these elements than the randomized controlled trial counterpart could. Alternatively, comparable information to that collected from a randomized controlled trial could be

obtained from an observational study at a cost saving. This however must be balanced with the fact that in a trial you can specify the variables of interest and obtain highly specific information on them, whereas routinely collected data is not intended for research and so does not allow such specification.

2.9. Motivation for using instrumental variable designs

Typically, new interventions are investigated using randomized controlled trials, whereby patients are assigned to either the treatment of interest or a placebo and then compared. In observational pharmacoepidemiology, drug interventions are investigated without assignment to treatment, using existing data sources. The former is considered the gold standard; however, it is not always ethical or practical to intervene. For example, to assess the effect of using antihypertensive treatments in midlife on later life risk of dementia would require a very long follow-up and consequently be very costly. For these reasons, I will be using observational data. However, there is an argument for emulating the randomized controlled design so that I can take advantage of its features, such as the concept of randomization. (73) Randomisation is used to minimize biases, such as those described in Section 2.8.1, as the random assignment ensures an equal balance of covariates between the two groups. This means any difference between the treatment and control groups can be attributed to the treatment. Instrumental variable analysis, which will be described fully in Chapter 3, estimates the average causal effect of an exposure on an outcome by considering their association with a third variable, known as the instrument or instrumental variable. The randomization assignment in a double-blind randomized controlled trial is an example of an instrument. In this thesis, I will use 'randomization-like' variables as instruments to analyse two forms of observational data: electronic health record data and genetic data. If these variables meet the instrument conditions provided in Section 3.2, then they should not be subject to the confounding and reverse causation issues that may have affected the observational studies conducted to date and can provide new evidence concerning repurposing antihypertensives for dementia prevention.

2.10. Summary

In this chapter, I introduced the public health problem that this research relates to – specifically, the continued lack of treatment options for the prevention of dementia despite the large numbers of people diagnosed with the condition each year. I then summarized the epidemiology of dementia and introduced the four existing treatments that can only address the symptoms of the condition. Further discussion of these treatments can be found in Chapter 6, which examines how these treatments are prescribed in England and the factors affecting their prescription. The chapter then moves onto discussion of drug repurposing and antihypertensive drugs as potential candidates for dementia prevention. This summarizes the existing evidence for this hypothesis, prior to the assessments conducted within this thesis. The final sections concern observational pharmacoepidemiology and my motivation for using instrumental variable designs to analyse such data. These sections include discussion of the strengths and limitations of observational pharmacoepidemiology, including that these methods can be implemented when randomized controlled trials are not feasible or ethical and they can have much larger sample sizes due to the reduction in cost and time required to implement them.

Chapter 3. Methods: Instrumental variable analysis using electronic health record data

3.1. Introduction

The analysis of electronic health records in this thesis focuses on instrumental variable estimation. Instrumental variable analysis estimates the average causal effect of an exposure on an outcome by considering their association with a third variable, known as the instrument or instrumental variable. This approach accounts for confounding of the exposure-outcome association, regardless of whether it was measured, and is often used when estimating unmeasurable exposure-outcome associations. (24) To be an instrument, a variable must meet three conditions, which are discussed in detail in Section 3.2.1. When met, these conditions allow the instrument to be used to identify bounds for the average causal effect. (74) This is known as partial identification of causal effects. A fourth assumption concerning the instrument-outcome association is required to obtain a point estimate of the average causal effect. Commonly used assumptions are effect homogeneity and monotonicity, which are covered in Section 3.2.2. (74) Together, the four assumptions allow the effect of the exposure on the outcome to be estimated using the instrument-outcome and instrument-exposure associations as follows:

$$\text{Effect of the exposure on the outcome} = \frac{\text{Instrument-outcome association}}{\text{Instrument-exposure association}}$$

The instrumental variable analysis model discussed here is often represented on a directed acyclic graph, as was previously demonstrated in Figure 1.1.

There are many instruments to choose from when conducting an instrumental variable analysis. This thesis will use physicians' prescribing preference as an instrument to proxy drug exposure using data from electronic health records. A description of this instrument can be found in Section 3.3. This thesis will also use Mendelian randomization. Mendelian randomization is a form of instrumental variable analysis that uses one or more genetic variants to proxy exposure and is discussed in Chapter 4.

The current chapter will introduce the instrument assumptions, discuss the use of physicians' prescribing preference as an instrument when using electronic health record data, and conclude with the development of a power calculator for instrumental variable analysis in pharmacoepidemiology. The motivation for the latter part of this chapter was the lack of literature concerning power calculations for studies in pharmacoepidemiology available at the start of my PhD, as most power calculators available until this point were developed for Mendelian randomization. I therefore derived and validated a power formula for studies using a single binary instrument to analyse the causal effect of a binary exposure on a continuous outcome. This work was completed under the supervision of Neil Davies, Frank Windmeijer, Stephen Burgess, and Richard Martin. The results have been published in the International Journal of Epidemiology. (86)

3.2. Instrument assumptions

3.2.1. The three instrument assumptions

To conduct a valid instrumental variable analysis, the instrument must satisfy the following three conditions:

IV1. The instrument must associate with the exposure

IV2. The instrument must only affect the outcome through the exposure

IV3. The instrument and the outcome must have no common causes

These conditions are known as relevance, the exclusion restriction and independence and, as mentioned in the introduction, allow the identification of bounds for the average causal effect. (23)

Relevance is the only instrument assumption of the three that is directly testable. This is done by regressing the exposure on the instrument to test whether they are associated. The magnitude of this association reflects the strength of the instrument for the instrumental variable analysis. In otherwise identical studies, stronger instruments will provide more power to detect effects and result in greater precision. The remaining instrument assumptions, namely the exclusion restriction and independence, cannot be directly tested empirically but are falsifiable. (74)

3.2.2. The fourth instrument assumption

As highlighted earlier, a fourth assumption, such as effect homogeneity or monotonicity, is required to refine the bounds for the average causal effect to a point estimate. Before I discuss the options for this assumption, I would like to introduce the following terms that will aid the explanation. Consider a binary instrument, Z, to proxy a binary exposure, X. In this setup, the population can be separated into four subpopulations, as shown in Table 3.1. These subpopulations are unlikely to be identifiable in the data as it is usually unknown what the individual would have done if they received the alternate assignment however, they are relevant to the discussion of the effect we will obtain under the different fourth assumptions. (74).

Table 3.1: Subpopulations in an instrumental variable analysis with a binary instrument and binary exposure.

	Exposure received when the instrument $Z = 0$	Exposure received when the instrument $Z = 1$
Always-takers	1	1
Never-takers	0	0
Compliers	0	1
Defiers	1	0

3.2.2.1. Effect homogeneity

The assumption of effect homogeneity holds if the effect of the exposure on the outcome is the same for all individuals. For example, consider testing a new drug for the treatment of hypertension. The homogenous treatment effect assumptions require that the treatment causes a specific change in systolic blood pressure, say a 10mmHg reduction, for all individuals regardless of all other factors that may affect systolic blood pressure, such as weight and alcohol intake. This is rarely plausible and so weaker versions of the assumption have been derived, such as the assumption of no effect modification. Consider a binary instrument to proxy a binary exposure. The no effect modification assumption states "the average causal effect on the additive scale is equal by levels of Z [the instrument] in both the treated and in the untreated". (74) This assumption can also be made using a multiplicative, rather than an additive, scale.

Hernán and Robins (2006) have shown that effect homogeneity is an implausible assumption for binary outcomes, such as those used in this thesis. This is because heterogenous effects will always exist for a binary outcome unless one of three situations occur. That is: the treatment causes the outcome in the entire population; the treatment prevents the outcome in the entire population; or the treatment has no effect on the outcome. (23) I will therefore use an alternative fourth point-identifying assumption, monotonicity, which is discussed below, for my analysis.

3.2.2.2. Monotonicity

The assumption of monotonicity holds if the exposure received when the instrument $Z = 1$ is greater than or equal to the exposure received when the instrument $Z = 0$ for all individuals. In terms of the subpopulations defined in Table 3.1, this means there are no defiers. When this is the case, the instrumental variable estimate represents the average causal effect of the exposure on the outcome among the compliers – known as the local average treatment effect. (75) This is because the weighted instrument-outcome association in each of the other subpopulations is now zero, as there are no defiers and neither the never-takers or the always-takers are influenced by the instrument. Monotonicity can be a more reasonable assumption than effect homogeneity in some circumstances – for example, when considering a binary outcome as in the analyses in this thesis.

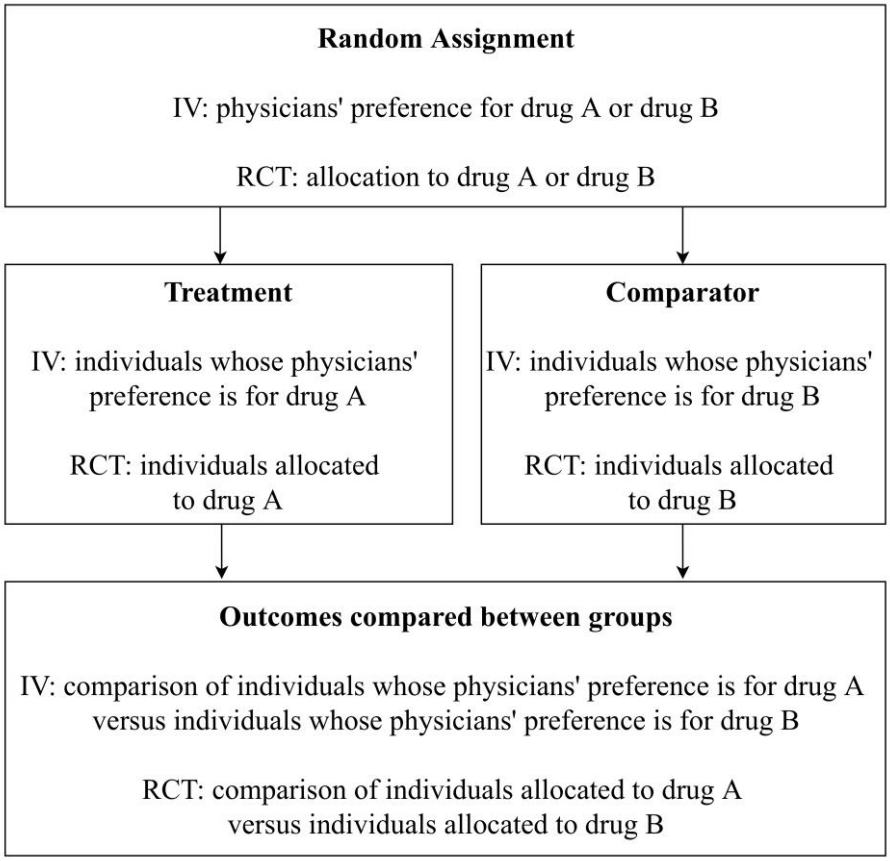
As for all assumptions, there are limitations to assuming monotonicity as the fourth instrument assumption. Firstly, it is not always plausible to assume that there are no defiers in an analysis. (74,75) In addition, there may be concerns regarding the relevance of an estimate that applies to compliers only. This is because compliers are rarely identifiable in the population and so it is difficult to target those for whom the estimate applies. Also, policy decisions apply to the whole population, which will include non-compliers, and so the estimate may not reflect the average treatment effect observed in the whole population.

3.3. Physicians' prescribing preference

Physicians' prescribing preference is a common choice of instrumental variable in pharmacoepidemiology. (27,28,76–79) It potentially meets the three instrument conditions as prescribing preference effects the prescription issued by the physician (IV1: relevance), it is unlikely to relate to the patient's outcome other than through the prescription issued (IV2: exclusion restriction) and it is not likely to share a cause with the patient's outcome (IV3: independence). A directed acyclic graph illustrating these assumptions in the context of the study presented in Chapter 7 is provided in Figure 7.2. A visual representation of how instrumental variable analysis using this instrument can be thought of as analogous to a randomized controlled trial is provided in Figure 3.1.

In this thesis, I will use physicians' prescribing preference defined by the past seven prescriptions issued by the physician as an instrument for exposure. This will give a score between zero and seven that indicates how many of the previous prescriptions have been in favour of the treatment of interest rather than the control treatment. I opted to use the instrument in a categorical form to improve instrument strength, however this instrument is often used in a binary form. For this, the instrument takes a value of one if the physician prescribed the treatment of interest more than the control treatment and a value of zero if the physician prescribed the control treatment more than the treatment of interest. Note that an odd number of prescriptions should be used to prevent ties. In the following section, I discuss the development of a power calculator for analyses using binary instruments such as this. I hope to develop the calculator further in the future so that it is applicable to analyses using categorical and continuous instruments however, in the meantime, it can be used to provide conservative power estimates for such analyses by dichotomizing the instrument.

Figure 3.1: Illustration showing how instrumental variable analysis using physicians' prescribing preference as an instrument can be thought of as analogous to a randomized controlled trial.



Naturally, physicians' will favour drug A or drug B and this will affect their prescribing. Instrumental variable analysis therefore uses this preference, which is unlikely to share a cause with the patient's outcome because patients have relatively little choice over which physician they see or knowledge of their physicians' preferences for certain drugs, to mimic allocation to drug A or drug B.

3.4. Power calculations

3.4.1. Motivation

Pharmacoepidemiological studies risk irrelevance if they are insufficiently powered to detect clinically meaningful treatment effects. Prior to starting a study, the statistical power to calculate a given treatment effect should ideally be calculated. This type of calculation is becoming increasingly important for grant and data request applications, which look to value the contribution of such studies, and is a requirement when applying for data from the CPRD – the main data source for this thesis (Chapter 5). However, when I applied for my CPRD data, there were not power calculators available for pharmacoepidemiological studies using instrumental variable analysis. This was the motivation for the final part of this chapter. Here, I present the derivation and validation of a power formula for studies using a single binary instrument to analyse the causal effect of a binary exposure on a continuous outcome in the context of pharmacoepidemiology. The code for this section is available from GitHub: https://github.com/venexia/PharmIV.

3.4.2. Formula derivation

The instrumental variable analysis under consideration requires the following three variables; namely a binary instrument Z, a binary exposure X and a continuous outcome Y. The outcome for patient i, for $i = 1, \dots, n$, is modelled as follows:

$$Y_i = \alpha + \beta X_i + U_i$$

where U_i is a zero-mean error term containing unobserved confounders, determining both the outcome Y_i and the treatment X_i. The instrument Z_i affects treatment X_i, but is not associated with the unobserved confounders and has no direct effect on the outcome.

Let $\tilde{Y}_i = Y_i - \bar{Y}$, $\tilde{X}_i = X_i - \bar{X}$ and $\tilde{Z}_i = Z_i - \bar{Z}$, where $\bar{Y}$, $\bar{X}$ and $\bar{Z}$ are sample averages. Denote by $\tilde{y}$, $\tilde{x}$ and $\tilde{z}$ the n-vectors of observations on $\tilde{Y}_i$, $\tilde{X}_i$ and $\tilde{Z}_i$ respectively. The two-stage least squares (2SLS) estimator of β is then given by:

$$\hat{\beta} = (\tilde{z}'\tilde{x})^{-1}\tilde{z}'\tilde{y}.$$

The variance of the 2SLS estimator is:

$$\mathrm{Var}(\hat{\beta}) = \sigma^2(\tilde{x}'P_{\tilde{z}}\tilde{x})^{-1}$$

where $P_{\tilde{z}} = \tilde{z}(\tilde{z}'\tilde{z})^{-1}\tilde{z}'$ and $\sigma^2 = E(U_i^2)$ is the residual variance. Note that conditional homoscedasticity holds so the variance is constant for all values of the instrument i.e. $E(U_i^2) = E(U_i^2|Z_i) = \sigma^2$ for $i = 1, \dots, n$.

Consider the term $\tilde{x}'P_{\tilde{z}}\tilde{x}$:

$$\tilde{x}'P_{\tilde{z}}\tilde{x} = \tilde{x}'\tilde{z}(\tilde{z}'\tilde{z})^{-1}\tilde{z}'\tilde{x} = n\left(\frac{\tilde{x}'\tilde{z}}{n}\right)\left(\frac{\tilde{z}'\tilde{z}}{n}\right)^{-1}\left(\frac{\tilde{z}'\tilde{x}}{n}\right)$$

Let $p_Z = P(Z = 1)$, $p_X = P(X = 1)$ and $p_{XZ} = P(X = 1|Z = 1)$. In large samples:

$$\left(\frac{\tilde{z}'\tilde{z}}{n}\right) \approx \mathrm{Var}(\tilde{z}) = p_Z(1 - p_Z)$$

$$\left(\frac{\tilde{x}'\tilde{z}}{n}\right) = \left(\frac{x'z}{n} - \overline{XZ}\right) \approx p_Z(p_{XZ} - p_X)$$

Hence $\tilde{x}'P_{\tilde{z}}\tilde{x}$ can be presented in the following way:

$$\tilde{x}'P_{\tilde{z}}\tilde{x} \approx \frac{n\big(p_Z(p_{XZ} - p_X)\big)^2}{p_Z(1 - p_Z)}$$

Now consider the instrumental variable estimator of β. Using the asymptotic distribution $\hat{\beta} \sim N(\beta, \sigma^2(\tilde{x}'P_{\tilde{z}}\tilde{x})^{-1})$, the distribution of the t-test statistic under the null hypothesis $H_0: \beta = \beta_0$ is:

$$t = \frac{\hat{\beta} - \beta_0}{\sigma\sqrt{(\tilde{x}'P_{\tilde{z}}\tilde{x})^{-1}}} \sim N(0,1)$$

The distribution of the test statistic under the alternative hypothesis $H_1: \beta = \beta_0 + \delta$ is:

$$t = \frac{\hat{\beta} - \beta_0}{\sigma\sqrt{(\tilde{x}'P_{\tilde{z}}\tilde{x})^{-1}}} = \frac{\hat{\beta} - \beta_0 - \delta}{\sigma\sqrt{(\tilde{x}'P_{\tilde{z}}\tilde{x})^{-1}}} + \frac{\delta}{\sigma\sqrt{(\tilde{x}'P_{\tilde{z}}\tilde{x})^{-1}}} \sim N\left(\frac{\delta}{\sigma\sqrt{(\tilde{x}'P_{\tilde{z}}\tilde{x})^{-1}}}, 1\right)$$

The null hypothesis is rejected if $|t| > c_\alpha$ where c_α is the critical value at significance level α.

The power is the probability the test statistic will exceed the critical value, which is:

$$P(t > c_\alpha) + P(t < -c_\alpha) = \Phi\left(-c_\alpha + \frac{\delta}{\sigma\sqrt{(\tilde{x}'P_{\tilde{z}}\tilde{x})^{-1}}}\right) + \Phi\left(-c_\alpha - \frac{\delta}{\sigma\sqrt{(\tilde{x}'P_{\tilde{z}}\tilde{x})^{-1}}}\right)$$

where $\Phi(s)$ is the cumulative standard normal distribution function evaluated at s. Power therefore increases as the value of σ decreases and/or the value of $\tilde{x}'P_{\tilde{z}}\tilde{x}$ increases.

By substituting $\tilde{x}'P_{\tilde{z}}\tilde{x}$ and simplifying, you obtain the following formula for power:

$$\text{Power} = \Phi\left(-c_\alpha + \frac{\delta(p_Z(p_{XZ} - p_X))\sqrt{n}}{\sigma\sqrt{p_Z(1-p_Z)}}\right) + \Phi\left(-c_\alpha - \frac{\delta(p_Z(p_{XZ} - p_X))\sqrt{n}}{\sigma\sqrt{p_Z(1-p_Z)}}\right)$$

The formula requires a total of seven parameters to be specified. This includes four parameters that must always be specified - these are the significance level, α; the size of the causal effect, δ; the residual variance, $\sigma^2 = E(U_i^2)$; and the sample size, n. As well as three that can be chosen from the following four parameters:

- The frequency of the instrument, $p_Z = P(Z = 1)$
- The frequency of exposure, $p_X = P(X = 1)$
- The probability of exposure given the instrument $Z = 1$, $p_{XZ} = P(X = 1|Z = 1)$
- The probability of exposure given the instrument $Z = 0$, $p_{XZ} = P(X = 1|Z = 0)$

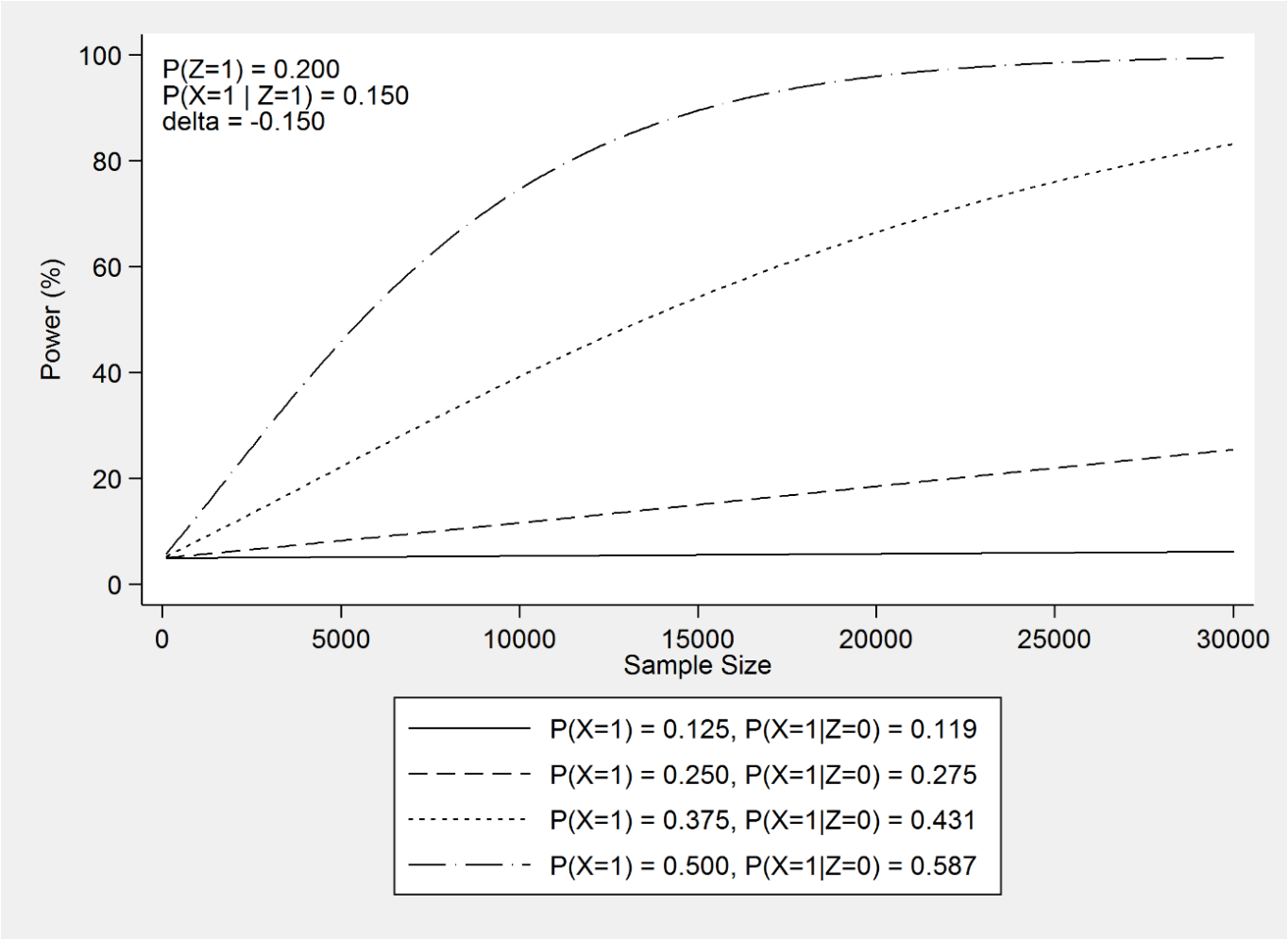
The chosen parameters must be specified so that the following holds:

$$P(X = 1) = P(X = 1|Z = 0)P(Z = 0) + P(X = 1|Z = 1)P(Z = 1)$$

The formula for power is available for use via an online calculator, which can be found at https://venexia.shinyapps.io/PharmIV/. The packages for R and Stata can be downloaded from https://github.com/venexia/PharmIV.

Note that the frequency of exposure in an instrumental variable analysis of this type is likely to be higher than a general population study because a drug is compared against one or more other drugs in a population of people with the indication for these treatments. General population studies on the other hand tend to compare a population who received the drug of interest with a population who did not receive it and consequently the frequency of exposure is generally much lower. The effect of varying the parameters within the formula on a study's power is best presented graphically. Figure 3.2 illustrates an example of the effect of the frequency of the exposure $p_X = P(X = 1)$ on the power of a study to detect a causal effect of $\delta = -0.150$ using an instrument with a frequency of $p_Z = 0.200$, a residual variance of $\sigma^2 = 1$ and a sample size of up to 30,000 participants. Both increasing the frequency of exposure up to 50% and increasing the sample size results in increased power for this study.

Figure 3.2: Illustration of the effect of varying the input parameters on the power formula I proposed.



The above power curves have been calculated for several values of the frequency of exposure $p_X = P(X = 1)$ *to show the effect on the power of a study to detect a causal effect of* $\delta = -0.150$ *using an instrument with a frequency of* $p_Z = 0.200$*, a residual variance of* $\sigma^2 = 1$ *and a sample size of up to 30,000 participants.*

3.4.3. Formula validation

To validate the power formula, I conducted a simulation by defining the three variables necessary to conduct instrumental variable analysis with a single instrumental variable as follows:

Instrument: $Z_i \sim \text{Binomial}(1, p_Z)$

Exposure: $X_i \sim \begin{cases} 0, \text{if } c_0 + Z_i(c_1 - c_0) + V_i \leq 0 \\ 1, \text{if } c_0 + Z_i(c_1 - c_0) + V_i > 0 \end{cases}$

Outcome: $Y_i \sim \delta X_i + U_i$

Where $p_Z = P(Z = 1)$ is the frequency of the instrument, $c_j = \Phi^{-1}\big(P(X = 1|Z = j)\big)$ for $j = 0,1$ are the inverse cumulative standard normal distribution, or quantile, functions of the conditional probabilities of exposure given the instrument, δ is the causal effect, and U_i and V_i are standard normally distributed error terms with covariance ρ.

The formula uses a binary instrument, binary exposure and continuous outcome and so the above variables were simulated to recreate data of this form. The instrument Z is modelled by a binomial distribution parameterised by its frequency $p_Z = P(Z = 1)$. This ensures a binary variable with the correct probability of success. The exposure X is also binary but is modelled using a threshold model. The variability in the equation for the exposure comes from the normally distributed error term V_i. The use of the model equation allows the exposure X to be associated with the instrument Z. The outcome Y is modelled by its model equation $Y_i = \delta X_i + U_i$. In the model, the instrument is valid as the outcome Y is only associated with the exposure X, as dictated by the causal effect δ, and is not associated with the instrument Z other than through the exposure X.

Using the generated data, I performed an instrumental variable analysis using the command ivreg2 in Stata. (80) From this analysis, I recorded the coefficient of the exposure X with the 95% confidence interval. I then counted the number of simulations for which the confidence interval excluded the null and divided this by the total number

of simulations to determine the power. By running the simulation and calculating the formula using the same parameters, I was able to validate the formula against the simulation.

The power calculated from both the simulation and the formula for several parameter combinations is presented in Table 3.2. The table contains 27 different simulations, and each was repeated 10,000 times. The simulations consider each combination of three values of the frequency of exposure, $p_X = 0.100, 0.250, 0.500$; three values of the probability of exposure given the instrument $Z = 1$, $p_{XZ} = 0.150, 0.300, 0.450$; and three values of the sample size, $N = 10000, 20000, 30000$. I set the frequency of the instrument, $p_Z = 0.200$; the causal effect, $\delta = -0.150$; the residual variance, $\sigma^2 = 1$; and calculated $P(X = 1|Z = 0)$ according to the following equation:

$$P(X = 1|Z = 0) = \frac{P(X = 1) - P(X = 1|Z = 1)P(Z = 1)}{1 - P(Z = 1)} = \frac{p_X - p_{XZ}p_Z}{1 - p_Z}$$

The formula and the simulation consistently provide similar results with an absolute mean difference of 0.4% for the parameter combinations presented. There is also no discernible pattern in the differences suggesting systematic bias is not present. Further to this, the power is consistent with its behaviour in other established power calculations. For example, increasing sample size universally improves power for all parameter combinations.

The effect of confounding was removed as a parameter from the formula because the power was insensitive to its value in the simulation setting. This is demonstrated in Table 3.3 that shows the simulation results for the same analysis with the effect of confounding taking different values between 0 and 1. Without accounting for the effect of confounding, the formula estimates the power of this analysis to be 32.3%. For the values of the effect of confounding listed, the simulation estimates power between 30.7% and 33.1%. As the power estimated by the formula lies within this interval, and the power estimated by the simulation is not monotonic for increasing values of the effect of confounding, this parameter was deemed unnecessary.

Table 3.2: Comparison of power as calculated by the power formula I proposed and by simulation.

p_X	p_{XZ}	10,000 patients		20,000 patients		30,000 patients	
		Formula	Simulation	Formula	Simulation	Formula	Simulation
0.100	0.150	6.6%	6.1%	8.3%	7.9%	10.0%	9.8%
	0.300	32.3%	33.3%	56.4%	55.5%	73.8%	73.9%
	0.450	74.7%	75.6%	96.0%	95.9%	99.5%	99.5%
0.250	0.150	11.7%	11.4%	18.6%	18.3%	25.5%	25.5%
	0.300	6.6%	5.4%	8.3%	7.9%	10.0%	9.8%
	0.450	32.3%	32.8%	56.4%	56.1%	73.8%	73.7%
0.500	0.150	74.7%	74.2%	96.0%	95.9%	99.5%	99.6%
	0.300	32.3%	32.5%	56.4%	57.1%	73.8%	73.7%
	0.450	6.6%	5.0%	8.3%	7.2%	10.0%	10.1%

The above results are based on an instrumental variable analysis where the significance level, $\alpha = 0.05$*; the size of the causal effect,* $\delta = -0.150$*; the residual variance,* $\sigma^2 = 1$*; and the frequency of the instrument,* $p_Z = 0.200$*.*

Table 3.3: The effect of confounding on power in the power formula simulation.

Effect of Confounding	Simulation
0.00	32.7%
0.10	33.1%
0.20	32.0%
0.30	31.8%
0.40	32.3%
0.50	31.9%
0.60	31.4%
0.70	30.7%
0.80	31.0%
0.90	31.3%
1.00	31.1%

The above results are based on an instrumental variable analysis where the significance level, $\alpha = 0.05$*; the size of the causal effect,* $\delta = -0.150$*; the residual variance,* $\sigma^2 = 1$*; the sample size,* $n = 10{,}000$*; the frequency of the instrument,* $p_Z = 0.200$*; the frequency of exposure,* $p_X = 0.100$*; and the probability of exposure given the instrument* $Z = 1$*,* $p_{XZ} = 0.300$*. The formula estimates the power of this analysis to be 32.3%.*

3.4.4. Limitations

As for any power calculation, you are limited by the formulae and parameters, which simplify the dataset being considered. Power calculated from such formulae cannot account for dataset characteristics outside of these parameters. For example, the formula makes no allowance for the presence of missing data – a known limiting factor on the power of a study. By allowing for missing data in the anticipated sample size, conservative estimates for the power of a study can be obtained using the formula presented. As mentioned earlier, further work is also needed in order to establish the formula for power in other scenarios that use instrumental variable analysis within a pharmacoepidemiology context. This includes analyses with binary outcomes and analyses that involve multiple instrumental variables.

3.5. Summary

In this chapter, I have introduced instrumental variable estimation and covered the assumptions necessary to conduct this type of analysis. I have also described the instrument that I will use for the analysis of electronic health records in this thesis, namely physicians' prescribing preference. The chapter concludes with the derivation and validation of a formula to calculate the power for instrumental variable analysis with a single binary instrument, binary exposure and continuous outcome in the context of pharmacoepidemiology. This work was motivated by my own application to the CPRD for data and has resulted in an online tool, as well as packages in R and Stata, that allow others to implement the formula. The next chapter will consider Mendelian randomization, which is a form of instrumental variable analysis that exploits genetic variation to proxy exposure, and is the other key method used in this thesis to assess whether antihypertensives can be repurposed for dementia prevention.

Chapter 4. Methods: Instrumental variable analysis using genetic data

4.1. Introduction

Mendelian randomization is a special case of instrumental variable analysis that assesses the causal effect of an exposure on an outcome by using one or more genetic variants to proxy the exposure. (81–84) The genetic variant(s) chosen to proxy an exposure will be a naturally occurring equivalent – for example, a drug that influences levels of an enzyme will be proxied by one or more genetic variants that determine the level of that enzyme in the body. Genetic variants meet the three instrument conditions, introduced in Section 3.2.1. This is because genetic variants are known to associate with the exposure (IV1: relevance) as they are selected from strong, replicated signals in genome wide association studies (GWASs) that search for the genetic variants associated with a given phenotype across the genome. Genetic variants are unlikely to affect both the outcome and the exposure other than through the same pathway (IV2: exclusion restriction) due to Mendel's second law of independent assortment, which suggests that variants for one trait will be inherited independently of variants for other traits. Finally, the factors determining a genetic variant are likely to be independent of the factors that determine the outcome (IV3: independence) because randomization at conception ensures that the environment should be equally distributed across variants. Methodological advances mean that Mendelian randomization studies can use summary data of genetic variant-exposure and genetic variant-outcome associations from separate GWAS – known as two-sample Mendelian randomization. (85,86) I will take advantage of this methodological advance, which is explained in detail later in this chapter (Section 4.4).

Mendelian randomization has been used to assess the relationship between a wide range of exposures and outcomes. However, there was relatively little discussion focussed on using Mendelian randomization for drug repurposing when I commenced my PhD. This chapter explains how Mendelian randomization can be used to predict drug repurposing opportunities, as well as exploring the strengths and limitations of using this method in this context. It is based on work that I led on, under the supervision of George Davey Smith, Neil Davies, and Richard Martin and has been published in the International

Journal of Epidemiology. (32) The published work considers Mendelian randomization for the prediction of both adverse drug events and beneficial drug repurposing opportunities (collectively referred to as 'unintended drug effects'); however, this chapter will focus on predicting drug repurposing opportunities.

4.2. Background

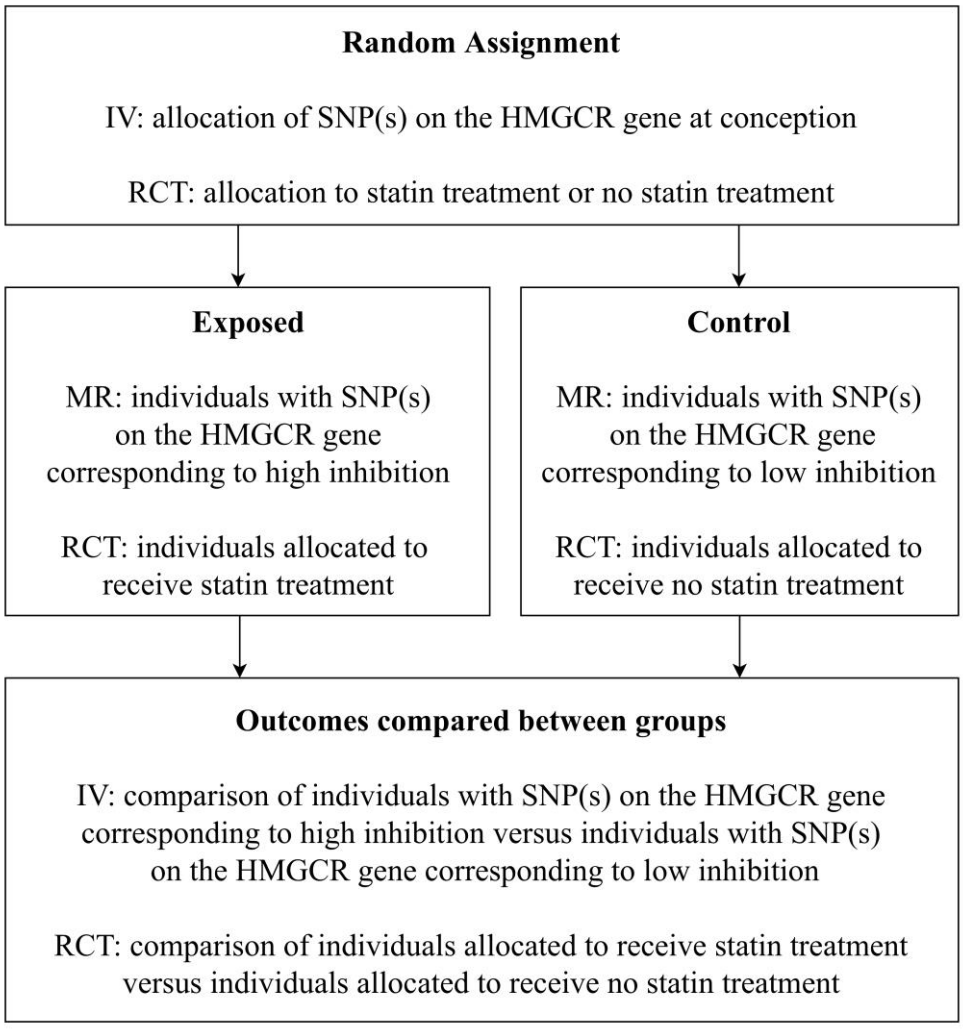
Drug repurposing opportunities have a desirable risk-versus-reward trade off, as alluded to in Section 2.5. This means they are often sought directly by pharmaceutical companies using purpose-built technology platforms. (4) Strong signals identified in these databases are then investigated using data from a range of sources, including randomized controlled trials either pre- or post-approval of the drug, meta-analyses of such trials, observational studies, and information from basic science. (53) However, these traditional approaches to predicting drug repurposing opportunities, particularly observational studies, are likely to be subject to several biases. Potential biases include an inability to determine causality, confounding by indication and other usually unobserved confounders. To overcome such limitations, I proposed Mendelian randomization (81–83) as a novel approach for the prediction of drug repurposing opportunities. The use of this method in this way is yet to be fully realised. This has been demonstrated by Finan et al, who mapped GWAS signals to a "set of genes encoding drug (and druggable) targets" to quantify how much of the genome may be considered "druggable". They identified 144 licensed drug targets as having a "discordant disease association and target indication considered to imply a potential repurposing opportunity". (87) In particular, the synthesis of evidence from Mendelian randomization with that from other sources, in the spirit of triangulation, could improve causal inferences of drug effects. (20) This is demonstrated by the choice of methods presented in this thesis.

To demonstrate the use of Mendelian randomization for the prediction of drug repurposing opportunities, consider a Mendelian randomization study designed to predict such opportunities associated with statins. Statins inhibit the enzyme 3-hydroxy-3-methylglutaryl-coA reductase (HMGCR) to lower LDL cholesterol and consequently reduce the risk of coronary heart disease. To proxy exposure to statins, a Mendelian randomization study would use one or more SNPs located near the HMGCR gene, i.e.

the mechanism of statins. This can be thought of as analogous to a randomized controlled trial if patients were randomized based on genetic variation, as demonstrated in Figure 4.1. The key distinction between Mendelian randomization and a randomized trial is that Mendelian randomization can be done using routine genotyping data, without the exposure of patients to the drug.

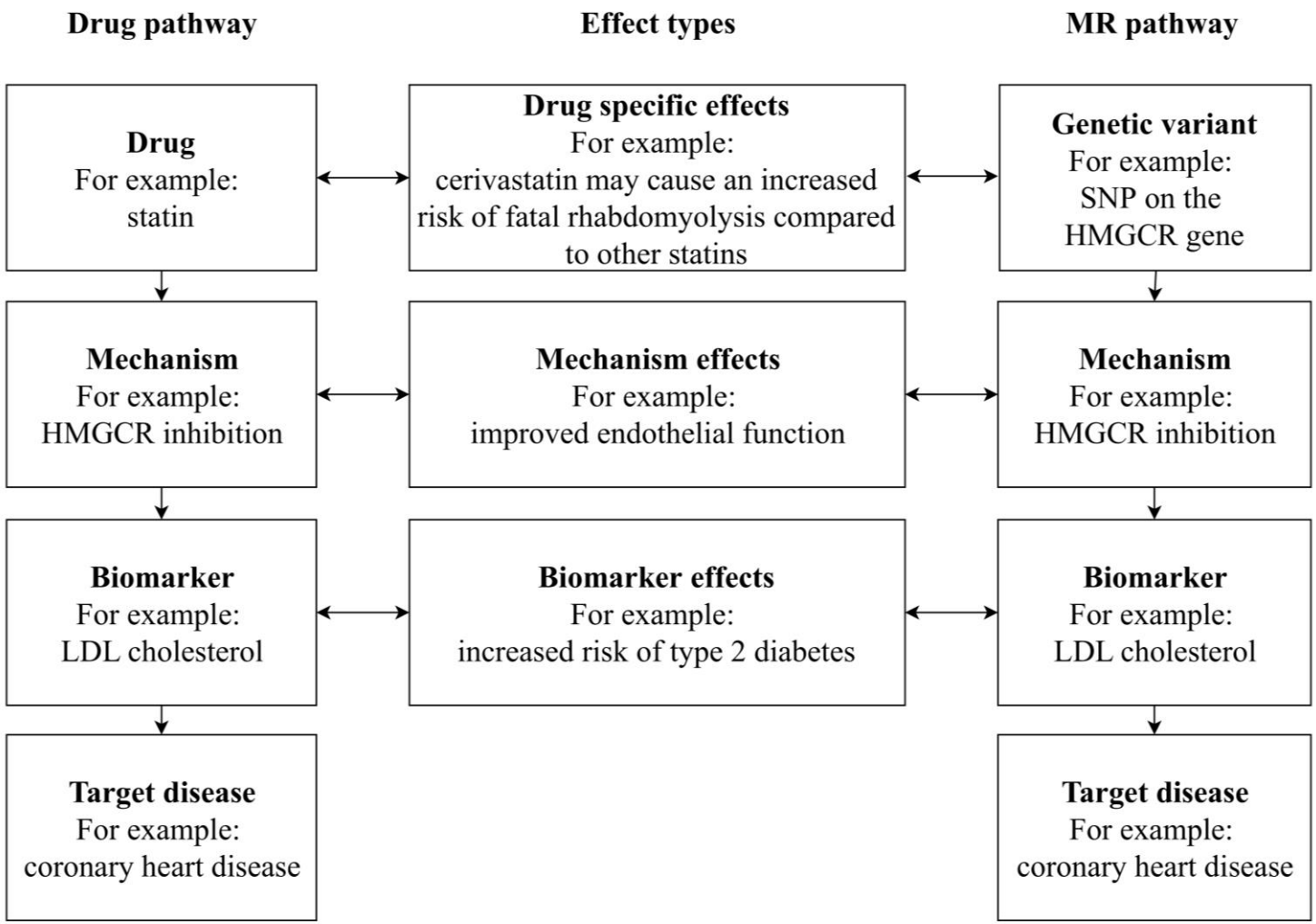
Potential drug repurposing opportunities can occur at different points along a drug pathway. This results in three different types of effect: drug substance specific effects; mechanism effects; and biomarker effects. These effects are presented in Figure 4.2 in terms of the statin example discussed previously. Note that to illustrate these effects in relation to a single drug class, the figure and the following explanation include adverse drug event examples. The first effect type, drug specific effects, arise because different compounds within the same drug class can have different effects. This is demonstrated in the case of statins by cerivastatin, which is suggested to have an increased risk of fatal rhabdomyolysis compared with other statins. This has led to it being withdrawn from the market. (88–92) The second effect type, mechanism effects, result from changes to a specific enzyme or biological pathway that are independent of changes to the biomarker. Several statins are thought to have lipid independent effects, i.e. changes resulting from HMGCR inhibition and not LDL cholesterol manipulation. These effects include improvement of endothelial function, though there is limited direct evidence for this in humans at present. (93–97) The final effect type, biomarker effects, are the effects arising from biomarker manipulation, which are independent of the mechanism of manipulation. For statins, these are effects resulting from changes in LDL cholesterol level, such as the increased risk of type 2 diabetes. (98–101) Ference et al demonstrated this effect by considering three mechanisms that lower LDL cholesterol: HMGCR, proprotein convertase subtilisin–kexin type 9 (PCSK9), and the LDL receptor. They found that "each set of gene-specific variants... had a very similar effect as the other sets on the risk of diabetes per unit decrease in the LDL cholesterol level." (102) Understanding the difference between these three effect types is key to understanding how Mendelian randomization can be used to predict drug repurposing opportunities.

Figure 4.1: Illustration showing how Mendelian randomization can be thought of as analogous to a randomized controlled trial.



To predict drug repurposing opportunities, the mechanism that the drug alters must be identified so that a suitable proxy for the drug can be found. Naturally, this mechanism will differ between individuals because of genetic variation. Mendelian randomization therefore uses the random allocation of genetic variants to mimic allocation (or not) to the drug of interest.

Figure 4.2: Illustration comparing when drug specific, mechanism, and biomarker effects occur in drug and Mendelian randomization pathways.



This diagram shows when different effects occur along the drug pathway and their equivalences in the Mendelian randomization pathway. Mendelian randomization can be used to predict mechanism and biomarker effects through the comparison of multiple mechanisms (Section 4.6.1.4; Figure 4.3). Genetic variants to proxy specific drug substances are rare and so this effect type is unlikely to be ascertained using Mendelian randomization.

4.3. Example

Yin et al investigated the relationship between serum calcium and the risk of migraine in a two sample Mendelian randomization study, with a genetic score that explained 1.25% of variation in serum calcium levels. They found "an elevation of serum calcium levels by a hypothetical 1mg/dl … was associated with an increase in risk of migraine (OR [odds ratio] = 1.80, 95% CI: 1.31 – 2.46, P =2.4×10^{-4})". (103) This result was supported by the two other methods implemented in the paper: an analysis of electronic health records and a genetic co-heritability analysis. Several therapeutic options were then proposed based on this evidence, including a drug named Cinacalcet. Cinacalcet is already approved by the Food and Drug Administration in the USA and works by antagonising the calcium-sensing receptor to lower calcium levels. The genetic variant rs1801725 is in the calcium-sensing receptor gene and mimics the action of the drug. As the genetic variant is associated both with serum calcium levels and increased migraine susceptibility, it highlights the potential to repurpose Cinacalcet for the treatment of migraine. However, the authors note that Cinacalcet is approved for the treatment of "secondary hyperparathyroidism or hypercalcaemia in patients with parathyroid carcinoma". Consequently, the use of this treatment for migraine has the potential to cause hypocalcaemia, among other side effects. In addition to Cinacalcet, the authors also highlight the potential to repurpose calcium channel blockers as a novel treatment for migraine. Existing evidence for this repurposing opportunity is mixed, however the authors suggest that the vasodilatory effects of calcium channel blockers, accompanied by direct manipulation of Ca^{2+} levels, could be beneficial based on their findings.

4.4. Two-sample Mendelian randomization

Two-sample Mendelian randomization uses summary data from different data sources for the instrument-exposure and instrument-outcome associations. These summary data are often obtained from databases of GWAS results, such as MR-Base (http://www.mrbase.org/). (104) Using two-sample Mendelian randomization has several advantages, most notably an increase in statistical power to detect causal effects. This increase in power is usually due to the increased sample size for the outcome, which often has a weaker association with the genetic variant chosen to proxy exposure and so limits the precision of the causal estimate obtained from Mendelian randomization. (105)

A further advantage of two-sample Mendelian randomization is that it is less likely to be affected by winner's curse. The curse occurs in one-sample Mendelian randomization when the instrument-exposure association is inflated, leading to its discovery (hence 'winners'), but also to an underestimate of the causal estimate obtained from Mendelian randomization. (106) Finally, two-sample Mendelian randomization, unlike one-sample Mendelian randomization, is biased towards the null in the presence of weak instruments. Further strengths and limitations of two-sample Mendelian randomization, compared to one-sample Mendelian randomization, are discussed in detail by Lawlor et al. (85)

4.5. MR-PheWAS

Mendelian randomization can either be used on a single outcome, as described above, or combined with a 'phenotype screen' for the prediction of previously unsuspected drug effects (beneficial or adverse) on a wide range of outcomes. Phenome wide association studies (PheWAS) are an example of a phenotype screen that search for phenotypes associated with a given genetic variant. This contrasts with GWAS, which search for the genetic variants associated with a given phenotype. MR-PheWAS were proposed by Millard et al and use "automated screening with genotypic instruments to screen for causal associations amongst any number of phenotypic outcomes". (107) This approach is hypothesis-free and could therefore be of great use for generating hypotheses concerning potential drug repurposing opportunities, particularly prior to the approval of a drug for a novel indication. Note that the hypothesis-free nature of this approach means that it will predict both beneficial and adverse effects, i.e. both drug repurposing opportunities and adverse drug events.

Limited phenotypic screening with Mendelian randomization has previously been demonstrated in the literature by the Interleukin 1 Genetics Consortium. (108) The consortium studied the long-term effects of interleukin 1 inhibition by combining the SNPs rs6743376 and rs11687782, identified in a GWAS they conducted, in a genetic score. (109) The score was then used to test the causal effect of interleukin 1 inhibition on rheumatoid arthritis and four cardiometabolic diseases. They found that "Human genetic data suggest that long-term dual IL-1α/β inhibition could increase cardiovascular risk and, conversely, reduce the risk of development of rheumatoid arthritis." (108) Note that

these results do not necessarily extend to inhibition of either of these interleukins alone. For example, a randomized controlled trial has found inhibition of IL-1β alone can reduce the risk of cardiovascular events. (110)

Since the work of the Interleukin 1 Genetics Consortium, the development of tools such as MR-Base has made the implementation of Mendelian randomization in this way much simpler. (104) For instance, MR-Base contains a database of harmonized GWAS summary statistics that greatly reduces the work required to test multiple outcomes for a phenotypic screen. The provision of such tools therefore means there is immense potential to use Mendelian randomization for the prediction of drug repurposing opportunities, with or without a priori hypotheses. Furthermore, there is the possibility of combining Mendelian randomization, and the related genetic methods, with non-genetic approaches in order to better explore the relationship between the genome and phenome. (111) This is discussed by Bush et al who consider the possibility of linking genetic data with that from electronic health records and epidemiological studies to better characterize "the impact of one or more genetic variants on the phenome" in the PheWAS setting. (112) This type of approach could be a particularly powerful tool for the prediction of drug repurposing opportunities, however, is not possible using the CPRD – the main data source for this thesis – at the time of writing.

4.6. Strengths and limitations

I will now highlight some of the strengths and limitations of Mendelian randomization, particularly those that make it suited to the prediction of drug repurposing opportunities and those that it may be susceptible to in this context. The strengths and limitations discussed are summarized in Table 4.1.

Table 4.1: Strengths and limitations of Mendelian randomization for drug repurposing.

Strengths	• Addresses confounding by indication • More robust to non-genetic confounding and reverse causation • Can be used either pre- or post-approval of a drug • Able to predict combined effects of drugs • Aids the distinction of mechanism and biomarker effects • Addresses missing data • Limits associative selection bias • Minimizes regression dilution bias
Limitations	• Rare effects may not be detected • Choice of genetic variant can lead to missed effects or conflicting results • Horizontal pleiotropy • Estimates are of lifelong exposure • Lack of genetic variants concerning disease progression • Collider bias in case-only studies • Genomic confounding • Weak instrument bias • Linkage disequilibrium • Combining genetic variants within a model can confound results

4.6.1. Strengths

4.6.1.1. Addresses confounding by indication

If the factors that determine whether a patient is exposed are also the factors that determine whether a patient experiences an outcome, observational studies can be subject to confounding by indication (Section 2.8.1.4). This is because an artificial association is induced between the exposure and the outcome. (113) Ultimately, these artificial associations can lead to incorrect inference being made concerning a potential drug repurposing opportunity. In Mendelian randomization, a genetic variant is used to proxy exposure. The genetic variant is unlikely to be affected by the indications for the exposure, so the possibility of confounding by indication affecting the results is reduced. This can be demonstrated by considering statins prescribed for the prevention of coronary heart disease. Statins are indicated for the treatment of existing cardiovascular disease. However, patients with this indication are at an increased risk of death. Consequently, there is an observational association between statin use and increased risk of cardiovascular death. However, this association is not due to statin use – it is an artefact of existing cardiovascular disease, an indication for statin use. (114) Mendelian

randomization reduces confounding by indication as the SNP located on the HMGCR gene, used to proxy exposure to statins, is a germline variant. This means it won't be a result of the existing cardiovascular disease and avoids confounding by indication.

4.6.1.2. More robust to non-genetic confounding and reverse causation

Mendelian randomization uses germline genetic variants that are less likely to be confounded by environmental, lifestyle or disease-related factors operating later in life. (84,115,116) Consequently, if a genetic variant is associated with an outcome only through its association with a drug effect, it is likely to be because the genetic variant causes the outcome. (117) Thus, Mendelian randomization should provide robust evidence about the causal effects of intervening on specific biological pathways. This is particularly important when considering physiological factors that change over the life course, such as LDL cholesterol and oestrogen levels, because the association of such factors with the outcome is likely to be heavily confounded by environment and lifestyle factors, as well as potentially being subject to reverse causation. For example, epidemiological studies have previously suggested that hormone replacement therapy could be protective against coronary heart disease. Results from these studies are summarized in a meta-analysis by Stampfer et al that found the relative risk to be 0.56 (95% CI 0.50-0.61). (118) However, these results are contrary to a number of clinical trials. (119) Lawlor et al suggest a possible explanation for this contradiction is the effect of early life socioeconomic position. They found "adverse socioeconomic factors from across the life course were associated with use of HRT [hormone replacement therapy]" in a study using data from the British Women's Heart and Health Study. (120) A Mendelian randomization study, which should not be subject to bias caused by socioeconomic position at any point in the life course, has since been conducted using data from young women in Hong Kong and older women in the Guangzhou Biobank Cohort Study. Unlike the observational studies, the Mendelian randomization analysis, which used genetically higher 17b-estradiol as a proxy for hormone replacement therapy, was in line with the results of the clinical trials. It concluded that "genetically higher 17b-estradiol was not associated with any cardiovascular disease-related risk factor or with Framingham score (0.01, 95% confidence interval = -1.34 to 1.31)." (121) This Mendelian randomization analysis therefore confirms that hormone replacement therapy

is unlikely to be a suitable drug repurposing candidate for coronary heart disease, without concerns about bias due to socioeconomic position.

4.6.1.3. Can be used either pre- or post-approval of a drug

As highlighted earlier, Mendelian randomization can be implemented using routine genotyping data. This means a study using this type of analysis will not require patients to be exposed to the drug and can be conducted at any point during the drug development process and beyond. This has several benefits including: improving trial safety through pre-specification of likely adverse outcomes, improving trial efficiency by identifying worthy drug targets, and reducing the possibility of exposing patients to unnecessary risks and harm. The benefit of implementing Mendelian randomization prior to a trial has recently been demonstrated by the Selenium and Vitamin E Cancer Prevention Trial, which found that selenium did not lower prostate cancer risk but did increase the risk of type 2 diabetes. Following the trial, a Mendelian randomization study found genetically elevated selenium was not associated with prostate cancer risk and was positively associated with type 2 diabetes risk – mirroring the results of the trial. (122) Implementation of Mendelian randomization prior to the trial could therefore have been an informative step in the assessment of selenium as a possible chemoprevention target.

4.6.1.4. Able to predict combined effects of drugs

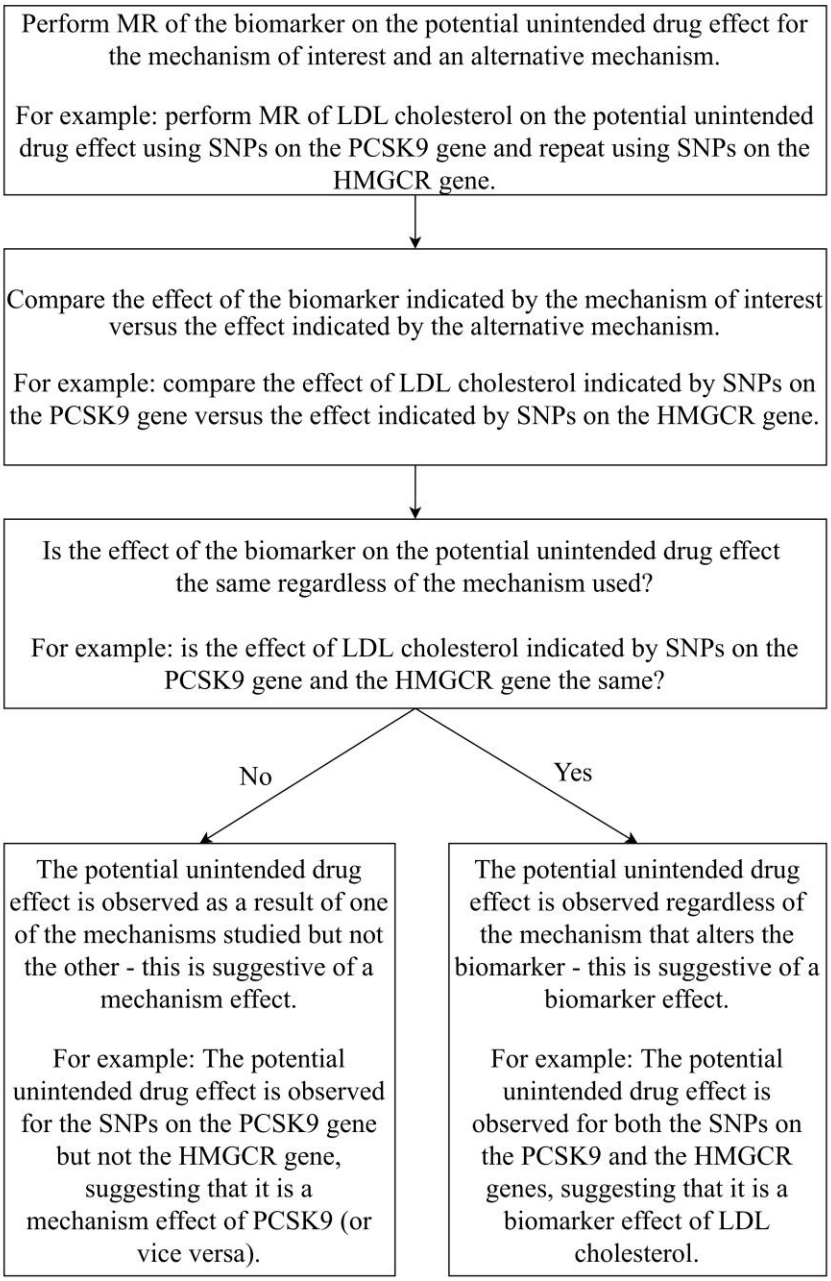
Mendelian randomization can be used to predict the combined effect of drugs by using a factorial design. This design separates patients into groups according to the first treatment and then, within those groups, separates patients into groups according to the second treatment and so on. In the case of two drugs (known as 2x2 factorial design), this results in four patient groups – one for each possible combination of two drugs. Assessing drug combinations is an important consideration as many medicines are only licensed for use when other treatments are either being used concurrently or have been previously used and failed. This makes the assessment of the 'additive' effect of drugs increasingly important. This approach is yet to be illustrated in the literature for a drug repurposing example, however it has been used to demonstrate the effect of combining an existing treatment to reduce LDL cholesterol, statins, with a novel treatment, PCSK9 inhibitors. The Mendelian randomization study was conducted by Ference et al and was published

prior to the corresponding trials. It found PCSK9 variants (the proxy for PCSK9 inhibitors) to have a similar effect as HMGCR variants (the proxy for statins) on the risk of cardiovascular events (OR 0.81, 0.74-0.89 vs OR 0.81, 0.72-0.90) and the risk of type 2 diabetes (OR 1.11, 1.04-1.19 vs 1.13, 1.06-1.20) for each 10 mg per decilitre decrease in LDL cholesterol level. (102) The Further Cardiovascular Outcomes Research With PCSK9 Inhibition in Subjects With Elevated Risk (FOURIER) trial was one of several PCSK9 inhibitor trials and considered the drug evolocumab for the reduction of LDL cholesterol in a population taking statins. It found the lipid lowering effect of evolocumab was in line with statins and "the rates of adjudicated cases of new-onset diabetes did not differ significantly between the two groups (hazard ratio, 1.05; 95% CI, 0.94 to 1.17)". (123) Both analyses agreed that the lipid lowering effect of the two drugs was additive. This was investigated by Ference et al by considering the combination of PCSK9 and HMGCR variants in a 2x2 factorial Mendelian randomization approach. Given the overlap of the Mendelian randomization and the FOURER trial, further trial data is required to assess the risk of type 2 diabetes associated with the use of PCSK9 inhibitors. However, the consistency of analyses demonstrates the value of Mendelian randomization, particularly when considering the combined effects of drugs prior to patient exposure. Such investigations are an important element of the safety assessment of a drug, even when considering the repurposing of a drug already approved for use, as the drug is likely to be taken alongside others when prescribed to the general population.

4.6.1.5. Aids the distinction of mechanism and biomarker effects

Mendelian randomization is able to distinguish mechanism and biomarker effects by enabling a formal statistical comparison of the effect of a biomarker influenced by different drug-related mechanisms, as illustrated in Figure 4.3. (101,124,125) As genetic variants are unlikely to proxy specific drug substances; drug substance specific effects are difficult to ascertain. Mechanism effects are observed when results differ between analyses that consider different mechanisms for the same target disease. Biomarker effects are observed when the results do not differ between these analyses.

Figure 4.3: Illustration of how Mendelian randomization can be used to distinguish mechanism and biomarker effects of drugs.



This diagram shows that if a potential unintended drug effect is indicated by the SNPs corresponding to multiple mechanisms then it is suggestive of a biomarker effect. This is because the effect occurs regardless of the mechanism used to induce the change. If this is not the case, the unintended drug effect is suggestive of a mechanism effect relating to the SNPs that indicated it. This is because the effect is specific to just one mechanism that induces a change in the biomarker and not all possible mechanisms.

In practice, this is investigated by considering multiple studies of the same pathway or combination analyses, such as that explored earlier by Ference et al. In the Ference et al example, it was suggested that the cause of increased type 2 diabetes risk may be related to an LDL receptor-mediated pathway, i.e. may be a biomarker effect. This was due to the similarity of the effects observed for HMGCR variants (the proxy for statins) and for PCSK9 variants (the proxy for PCSK9 inhibitors). Further evidence for this hypothesis has been gained through specific assessment of potential shared pathways by Ference et al and other genetic studies that have found similar results. (101,102,126)

4.6.1.6. Addresses missing data

As explained in Section 4.4, two-sample Mendelian randomization is implemented using summary data about genetic variants meaning it does not require individual level data. In theory, provided there are robust genetic variants for the drug exposure of interest and there is a large GWAS of the outcome, Mendelian randomization should not be limited by missing or incomplete data. This could potentially overcome an issue that is often problematic in other observational study designs. (127) However, this has not been formally assessed to date and warrants further research.

4.6.1.7. Limits associative selection bias

Associative selection bias occurs when patients are selected into a study due to both their exposure and disease statuses. (82) This is also known as Berkson's bias, named after Joseph Berkson, and can be demonstrated by considering a scenario where patients are selected from a hospital. (128) In this case, the fact that the patient is in hospital effects both their exposure and disease statuses and consequently induces bias. This is because the selection of patients from the hospital means the analysis is conditioned on a binary variable that indicates whether a patient is in hospital or not and only the sub population for which the variable 'in hospital' is true have been analysed. (129) Such bias is reduced in Mendelian randomization as the factors that led patients to have the value true for the variable 'in hospital' are unlikely to be the factors that determined the genetic variants they received from their parents at conception. This prevents association of the instrument for exposure with the outcome through these confounding factors.

4.6.1.8. Minimizes regression dilution bias

Mendelian randomization estimates are of lifelong exposure, which can mean estimates do not match those obtained from clinical practice. This is discussed in detail later as a limitation of Mendelian randomization. However, there are some benefits associated with this, specifically that it minimizes regression dilution bias. Regression dilution bias occurs as a result of measurement error in the exposure that, at the population level, can lead to an underestimate of the causal effect of an exposure on an outcome. As Mendelian randomization uses genetic variants and measures lifelong exposure, attenuation due to measurement error in the exposure is much less likely to be an issue. (82)

4.6.2. Limitations

4.6.2.1. Drug repurposing opportunities for rare diseases may not be detected

Single sample Mendelian randomization studies, where the instrument-drug and instrument-outcome associations are recorded in the same dataset, are not suited to detecting drug repurposing opportunities for rare diseases. This is because the data for these diseases, if available, is likely to lack the power needed to identify an effect. This has been demonstrated in the literature by a rare but serious potential adverse effect of statins: rhabdomyolysis. The global incidence of rhabdomyolysis is unknown but it is uncommon, with an estimated 26,000 cases per year occurring in the US according to the 1995 National Hospital Discharge Survey. (130,131) Prevalence estimates of the condition are also difficult to obtain – this is due in part due to the previously disputed clinical definition for the condition. (132) The rare nature of rhabdomyolysis means that Mendelian randomization studies drawing on information from one dataset, as is the case for single sample Mendelian randomization, are unlikely to have sufficient power to detect it as a mechanism effect of statins. (133–136) For example, in UK Biobank (one of the largest studies with genetic data available at present), just 322 of 392,242 participants with ICD-10 diagnosis codes have the M62.8 code for ‘other specified disorders of muscles’, which includes rhabdomyolysis. This number would be further reduced once participants were restricted to those with statin exposure prior to the onset of the condition. This means you cannot use single-sample Mendelian randomization to

disentangle whether rhabdomyolysis is a mechanism effect of statins that is more pronounced for cerivastatin or a drug-specific effect of cerivastatin. Two-sample Mendelian randomization studies, where the instrument-drug and instrument-outcome associations are recorded in separate datasets, may be the best approach to overcome this limitation. This is because the two-sample approach allows the use of a case-control GWAS for the rare event, which is likely to provide a greater number of cases than other study designs. Predicting drug repurposing opportunities for rare diseases will require a hypothesis free approach and, while this is theoretically possible, it will be hard to achieve with the currently available resources. To make this approach feasible, databases of GWAS for classical rare unintended drugs effects - including GWAS implementing case-control designs - should be curated so that drugs can be tested against them in a Mendelian randomization framework.

4.6.2.2. Choice of genetic variant can lead to missed or conflicting effects

Drug repurposing opportunities may be missed if you chose a genetic variant to proxy exposure downstream of the effect you are interested in. Consider once more the statin example used to illustrate the effect types that can be investigated using Mendelian randomization. Using genetic variants at the biomarker level, i.e. those related to LDL cholesterol level, to investigate statins will miss the mechanism effects, such as the lipid-independent improvement to endothelial function. Alternatively, the choice of genetic variant may lead to conflicting results. This can happen if the chosen genetic variants alter the relationship between the exposure and the biomarker or effect multiple biomarkers related to a single disease. Given these issues, careful consideration must be given to the choice of genetic variant used for Mendelian randomization analyses.

4.6.2.3. Horizontal pleiotropy

Horizontal pleiotropy occurs when a genetic variant influences multiple phenotypes through distinct pathways. (83) This can bias a Mendelian randomization estimate as the estimate will include the effect of the variant through pathways other than that of interest, violating the exclusion restriction assumption (IV2 – Section 3.2.1). Horizontal pleiotropy is opposed to vertical pleiotropy, which occurs when the instrument SNP(s) are associated with other phenotypes that occur between exposure and outcome or after

the outcome of interest and does not invalidate the Mendelian randomization assumptions. Methods such as weighted median and MR-Egger regression are potentially more robust to horizontal pleiotropy. (133,137) These methods use multiple SNPs and so estimates should not be as affected by SNPs subject to horizontal pleiotropy. In the case of the weighted median estimates, up to 50 per cent of SNPs included can be invalid instruments and the estimate will remain consistent. (133) Horizontal pleiotropy can also be assessed using leave-one-out analyses that calculate the estimate without each SNP in turn to identify whether a SNP is having an undue effect on the overall estimate. Using such methods and having the biological knowledge to support your chosen genetic variants, should help to minimize issues such as horizontal pleiotropy.

4.6.2.4. Estimates are of lifelong exposure

Mendelian randomization estimates indicate lifelong perturbations in an exposure. Therefore, careful consideration of the exposure and its timing must be made to avoid misinterpretation of results. (138,139) For example, some exposures are cumulative whereby repeated exposure, over a sustained period, results in the outcome. Mendelian randomization analyses of such exposures are likely to overestimate the effect observed in other study designs, including randomized controlled trials, as these designs consider much shorter periods of exposure with lower compliance. A further example is time-dependent exposures. Mendelian randomization analyses of this type of exposure can provide misleading evidence about the effect of manipulating an exposure after the critical period. This is because the Mendelian randomization estimate will, by definition, include any critical periods in its assessment of lifelong exposure.

4.6.2.5. Lack of genetic variants concerning disease progression

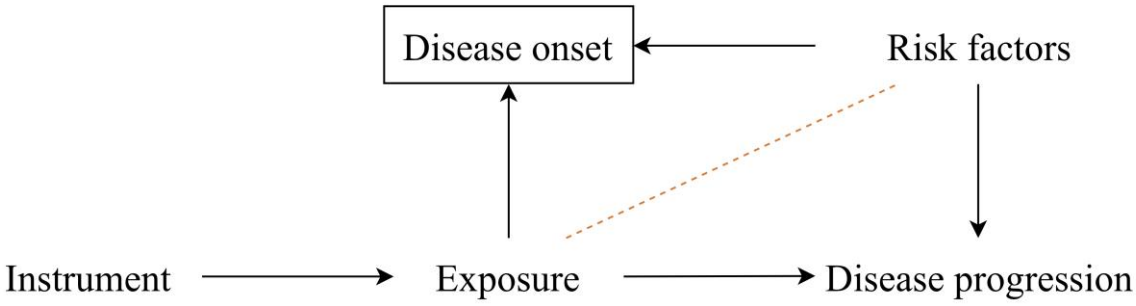
A large proportion of the genetic variants that have been identified to date are concerned with the incidence of disease. Consequently, it is difficult to use the proposed approach to predict drug repurposing opportunities that will alter the progression of a disease. In 2017, Paternoster et al reported that "only a small proportion of GWAS studies (~8% of associations curated in the GWAS Catalog ($p<1x10^{-5}$)) have attempted to identify variants associated with disease progression or severity and those that have are mostly small (90% have n<5000)". (140) To predict treatments that will alter the progression of a

disease in the future, there must be an increased focus on large GWAS concerning disease progression in the interim to fill this gap in the literature.

4.6.2.6. Collider bias in case-only studies

If a Mendelian randomization study uses only the cases of the disease - for example, when studying disease progression - it can be affected by collider bias (Section 2.8.1.2). (65) This occurs because the analysis is conditioned on disease onset and so an association is induced between the instrument and the confounders that effect both disease onset and progression. (140) This is illustrated in Figure 4.4. Overcoming this issue is an area of active research. (141)

Figure 4.4: Illustration of how collider bias can affect Mendelian randomization when only disease cases are analysed.



In this diagram, conditioning on disease onset induces an association between the exposure and the risk factors meaning the exposure can affect disease progression via an indirect path. Reproduced from Paternoster, L., Tilling, K., & Davey Smith, G. (2017). Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. PLoS Genetics, 13(10), e1006944. https://doi.org/10.1371/journal.pgen.1006944

4.6.2.7. Genomic confounding

Mendelian randomization can be subject to genomic confounding, which occurs when the causality of a genetic variant is misinterpreted. An example of this is population stratification. Genetic variants occur at different frequencies in different populations. This means if, for instance, different ethnicities have different rates of outcomes, differences due to ethnicity could be incorrectly ascribed to the risk factor of interest.

4.6.2.8. Weak instrument bias

Instruments are termed weak when the correlation between the instruments and the exposure is low. (142) A commonly cited threshold is a partial F statistic of the association between the instrument and the exposure of less than 10. (30,74) Weak instruments will result in low power to detect a causal effect. (143–145) They also induce bias, as such instruments may explain only a small proportion of the association between the exposure and outcome. Weak instrument bias is not specific to Mendelian randomization, however Mendelian randomization is particularly susceptible to it. This is because GWASs often investigate only common genetic variants or combine the effect of rare genetic variants. Consequently, individual genetic variants may explain very little of the observed variation.

4.6.2.9. Linkage disequilibrium

SNPs that have non-random associations with each other are said to be in linkage disequilibrium. (146) Many Mendelian randomization methods require SNPs not to be in linkage disequilibrium and so measures, such as clumping, must be taken in order to identify independent SNPs. This can limit the number of SNPs included in the final analysis. If SNPs in linkage disequilibrium are included in the analysis, confounding may be introduced through the non-random associations that exist between the SNPs. Ultimately, this could lead to incorrect inference concerning drug repurposing opportunities and negates the benefits of this method over non-genetic observational methods. SNPs in linkage disequilibrium are primarily an issue for two-sample Mendelian randomization, where the non-random associations between them lead to SNPs in high linkage disequilibrium regions being given excess weight in an analysis. One-sample Mendelian randomization using allele scores will not be affected by this issue.

4.6.2.10. Combining genetic variants within a model can confound results

The combination of genetic variants, especially when little is known about the biological effect of those variants, can confound estimates. For example, in Mendelian randomization studies considering drug repurposing opportunities for the progression of

disease, combining genetic variants can exacerbate the effect of collider bias. This is because a set of genetic variants associated with disease incidence will be inversely associated with any other set of genetic variants associated with disease incidence via another pathway. The combination of these sets will then bias the association of the genetic variants with progression. (140)

4.7. Summary

In this chapter, I have introduced Mendelian randomization as a method for predicting drug repurposing opportunities and covered the strengths and limitations of using the method in this way. I have also provided examples of Mendelian randomization predicting drug repurposing opportunities from the literature and discussed MR-PheWAS, an extension of Mendelian randomization that is particularly beneficial in this context. This chapter motivates the use of Mendelian randomization in Chapter 8, where I implement this method to assess whether antihypertensives can be repurposed for the prevention of dementia.

Chapter 5. Methods: The Clinical Practice Research Datalink

5.1. Introduction

This thesis makes use of the CPRD, an ongoing UK-based primary care database, established by Value Added Medical Products (VAMP) Health in 1987. VAMP Health incentivised clinical practices to join the database by offering free computer systems with ongoing maintenance in exchange for anonymised patient data. In 1993, the Department of Health acquired the database and it became known as the General Practice Research Database (GPRD). In 2012, the database received its current name, the CPRD. At present, the National Institute for Health Research and Medicines and Healthcare Products Regulatory Agency fund the database however other groups, such as the Wellcome Trust and Medical Research Council, fund studies using the data. (147) The CPRD's Independent Scientific Advisory Committee (ISAC) monitor all data requests for such studies and seek approval from an ethics committee, scientific committee and the National Information Governance Board Ethics and Confidentiality Committee when needed. (148) The protocol [ISAC Protocol 15_246R] for the data used in this thesis was submitted for a larger project that includes investigating the effects of commonly prescribed drugs on the prevention and treatment of dementia, Parkinson's disease and amyotrophic lateral sclerosis. These additional exposures and outcomes are discussed in this chapter to fully explain the data extraction process however only those relating to hypertension and dementia were ultimately used in this thesis. The protocol was published in BMJ Open prior to the commencement of the studies it describes. (36)

This chapter introduces the CPRD as a data source, including its strengths and limitations, and describes the structure of the data. I then introduce the additional data, available through data linkage, that is utilized in this thesis and provide more specific details about the study population I used in my analyses. With this context in place, I describe the steps I took to clean the data and define the events and diagnoses. The final section of this chapter covers the covariates I considered. All code lists referenced in this chapter have been made available online: http://rebrand.ly/repurposing-antihypertensives-dementia-codelists. (149,150) As has the code used for data cleaning: https://github.com/venexia/CleanCPRD. The information presented here is relevant to Chapters 6 and 7, both of which use the CPRD data.

5.2. Data source

The CPRD contains "over 11.3 million patients from 674 practices in the UK… and includes over 79 million person-years of follow-up". (147) Specifically, for each patient, the database contains all contact between them and their general practice. This means there are data available "on demographic information, prescription details, clinical events (symptoms, diagnoses), preventive care provided, tests, immunisations, specialist referrals, hospital admissions and their major outcomes, and details relating to death". (147) These data are primarily recorded by general practice staff either during a consultation or as a result of feedback from other sources, such as secondary care, using a system of codes (see Section 5.7). In addition to this, 58% of practices included in the CPRD have given permission for their data to be linked with other data sources, such as Office of National Statistics (ONS) and Hospital Episodes Statistics (HES), allowing further information concerning their patients to be obtained.

The CPRD is "broadly representative of the UK general population" with approximately 7% of the population registered in a participating practice in 2013. (147) At the last census in 2011, the CPRD was generally comparable for age, sex and ethnicity with a slight underrepresentation of younger people. (147,151) The database has also been shown to be representative of body mass index when compared against the Health Survey for England. (152) However, at a practice level, the CPRD is less representative with respect to the size and location of practices. (153) A 2010 systematic review found the CPRD to be accurate in terms of diagnostic coding, though acute conditions are not recorded as well as chronic conditions. (154) Of particular relevance to this thesis, the validity of codes for dementia diagnoses is generally reported to be high. (155)

5.2.1. Strengths of the data source

The CPRD has many strengths as a data source. (147,156) Firstly, it provides a large quantity of population-based data. As noted in Section 2.8.2.4, this provides greater statistical power to detect treatment effects and means the database can be used to study rare outcomes. Secondly, it includes access to outpatient information and original medical records, which may enhance the more conventional recorded data. Thirdly, as noted above, the database is representative of the UK population allowing researchers to

measure the effectiveness of treatments as opposed to their efficacy (see Section 2.8.2.1). Fourthly, the long follow up of patients means that outcomes with long latency can be studied (see Section 2.8.2.2). A further strength of the database is that it can be used when clinical trials are not ethical or practical, such as for studies concerning outcomes in pregnant women. Finally, it includes high quality measurement of some data such as those listed in the Quality and Outcomes Framework, for which data collection is incentivised. Historically, this has been shown to improve data collection coverage. (147)

5.2.2. Limitations of the data source

Despite the many strengths of the CPRD, the database is also subject to some weaknesses. (147,156) Most notably, the database has complex missingness that can be difficult to quantify. This issue is exacerbated by the fact that the absence of a code indicating a disease or prescription can reflect either genuine absence of that disease or prescription or a missing data item. Secondly, there is a lack of standardized definitions for diagnoses that may limit the comparability of studies. Researchers continue to work to overcome these issues as demonstrated by the setup of code list repositories to aid the standardization of definitions. (157) Thirdly, the core CPRD dataset mostly captures information from primary care and is therefore reliant on information from other services, such as secondary care, being manually added to records. Of particular concern is diagnoses received from a specialist that may not necessarily be re-recorded in primary care records. Fourthly, while the CPRD is generally representative of the UK population, there are several patient groups – for instance, patients at some residential homes – that are known to be missing. (147) Finally, the data included in the CPRD are electronic medical records and so are a secondary data source. This means data was not collected for a specific research question and you cannot specify how or what data is collected.

5.2.3. Rationale for using the data source

There are two key reasons why I chose to use the CPRD as the main data source for this thesis. The first key reason was the long follow-up of patients in the dataset. This was essential for my investigations into whether antihypertensive drugs have a causal effect on incident dementia as the drugs are typically prescribed in mid-life, while incident dementia typically occurs in late life. This meant a long follow-up was necessary to

ensure the outcome of interest was measured. Non-electronic-health-record data sources, such as clinical trial data, are unlikely to have the required length of follow-up for this type of analysis. The second key reason is the size and depth of the dataset. The CPRD is "one of the largest databases of longitudinal medical records from primary care in the world" and has linkages to other data sources such as ONS and HES. (147) By using one of the largest datasets possible, I maximized the size of the cohort for my study and consequently the power, allowing detection of even very small effects of my drugs of interest. By using data with linkages to other data sources, I was able to perform sensitivity and specificity analyses and obtain additional information on patients that was not present in the main dataset.

5.3. Data structure

The CPRD provide their data in tab-delimited format; arranged according to its type. There are ten distinct types of file:

- Patient
- Practice
- Staff
- Consultation
- Clinical
- Additional clinical details
- Referral
- Immunisation
- Test
- Therapy

Each patient in the database has a unique patient identifier that allows extraction of their details. Similarly, each practice and staff member have unique identifiers. The basic demographics and registration of the patients is contained in the patient file, the details of each of the practices in the practice file and the staff details in the staff file. This thesis will predominantly make use of the following five file types: clinical, referral, immunisation, test and therapy, as these are the files necessary to identify the diagnostic and treatment events of interest. The additional clinical details and consultation file will

support this information allowing access to information about consultation type and test results.

5.4. Data linkage

As part of the protocol, access to three linked datasets was requested. Firstly, the ONS death registry because cause-specific mortality is more accurately recorded in the registry than in the general practices that contribute to the CPRD. (158) Secondly, the HES data to allow investigations into service use, though this was not ultimately studied in this thesis. Finally, the Index of Multiple Deprivation, which provides a practice postcode-based indicator that can be used as a proxy to adjust for socioeconomic position.

5.5. Data specification

The data extract used in this thesis is taken from the March 2016 CPRD Gold snapshot and contains records from 1st January 1987 to February 29th 2016. To be included in the extract, patients had to qualify for one of the cohorts presented in the study protocol. (36) These were based on the diagnosis definitions provided in Table 5.1. In addition to having a diagnosis of interest: patients had to be aged 40 and over; attend an 'up to standard' practice; and have data deemed 'acceptable' by the CPRD.

5.6. Data cleaning

To use the data, I converted the tab-delimited files to Stata dataset files. During this process, I could format the variables containing dates so that they were ready for use in Stata functions. I also took the opportunity to compress the datasets to make the use of these files as efficient as possible in the analysis. I saved the resulting Stata dataset files in a directory labelled 'raw'. To use the data, I load from this directory and save into a working data directory to prevent any changes to the original data.

5.7. Defining events

There are two types of event relevant to this thesis, namely diagnostic and treatment events. Diagnostic events are defined using Read codes and test results, while treatment

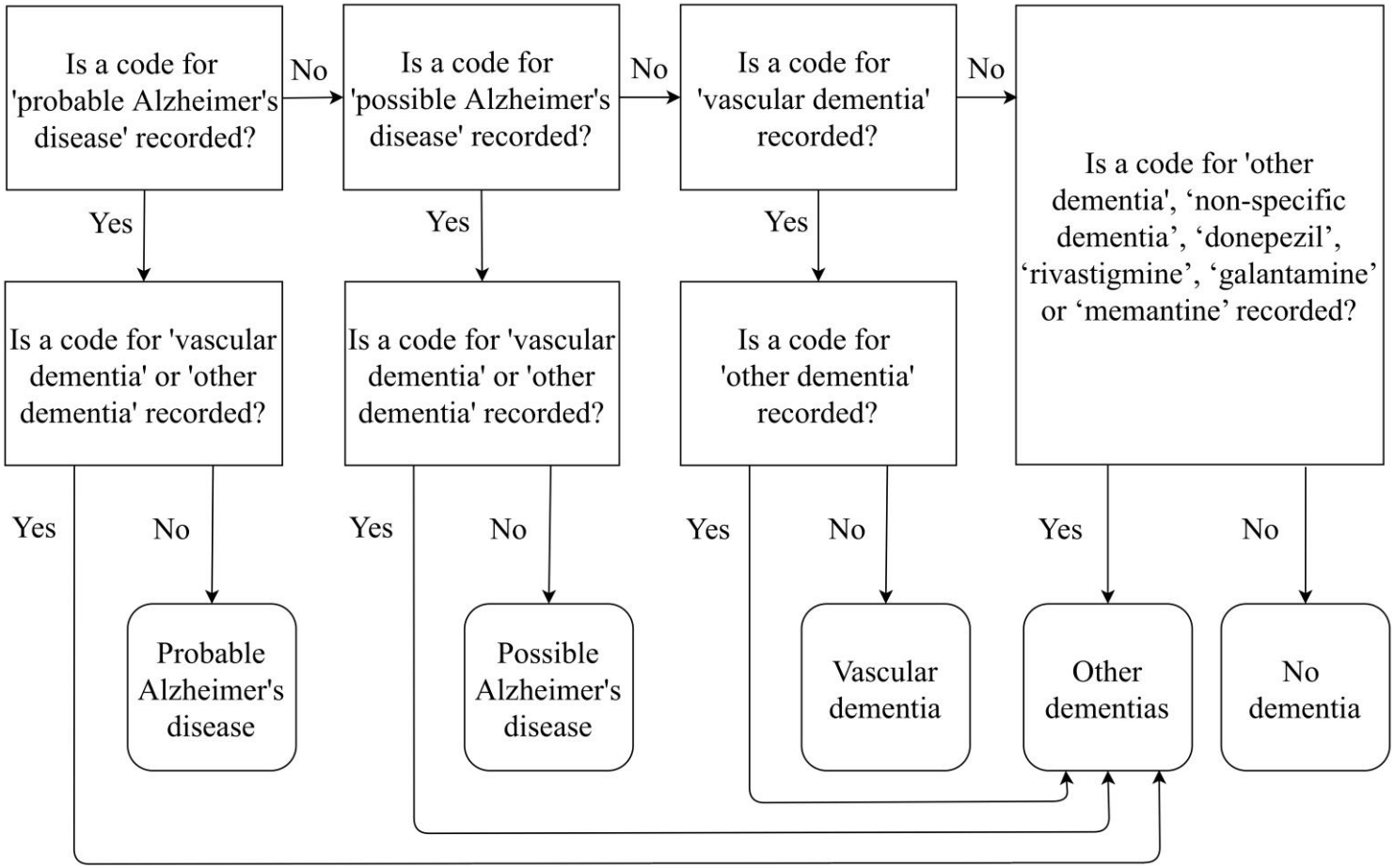
events are defined using product codes. Read codes and product codes uniquely identify clinical terms and prescriptions respectively in the CPRD. (159) Note that the linked data from the ONS and HES databases uses International Classification of Diseases (ICD) codes, not Read codes, and so equivalent code lists are required (Section 5.9).
To define events according to Read or product codes, I created code lists that detailed each of the codes relating to a given event. To determine relevant Read codes, I searched their descriptions using the medical dictionary within the CPRD code browser application (version 3.0.0). To identify the product codes for a given treatment, I searched the drug substance names provided by the British National Formulary guidance for each treatment in the product dictionary of the same application. Search terms for both the Read and product codes are provided in Appendix B. The code lists were later refined under the guidance of my supervisors to ensure only relevant codes were included and are available from the data.bris Research Data Repository. (150) For each code list, I matched patients in the relevant CPRD files with the listed codes using the 'joinby' command in Stata. Note that Read codes are recorded in four file types; namely clinical, referral, immunisation and test files; while product codes are recorded only in therapy files. Restricting the dataset to just the patients who matched a code on the code list allowed me to create an event list detailing the patient ID, the specific code recorded and the date. As the event list details each unique event, patients can appear more than once on this list. I therefore created a patient list for each code list that contained each patient only once with the date of their first recorded code and the number of entries they have on the event list. This assumes that the event date is the first date on which a relevant code is recorded.

To define events using test results, I took a similar approach in defining event and patient lists. The CPRD record test results in test files with an 'enttype' number and several columns containing data related to the test. For example, an 'enttype' number of one indicates a blood pressure reading and the diastolic and systolic blood pressure readings are recorded alongside this in data columns 1 and 2 respectively. The additional clinical details files do not contain event dates but instead provide an id, known as 'adid', which links to the clinical files. To simplify the extraction of additional clinical details, I created dated additional clinical detail files by merging the dates from the clinical files using the patient ID and the additional details ID. I was then able to extract the test results that

Table 5.1: Diagnosis definitions for analyses using the CPRD in this thesis.

Diagnosis	Definition
'At risk of' hypertension	Patients with one or more codes on any of the following lists: • At risk of hypertension • Systolic blood pressure test between 120-139 mmHg • Diastolic blood pressure test between 80-89 mmHg
Hypertension	Patients with one or more codes on any of the following lists: • Hypertension • Systolic blood pressure test of 140 mmHg or above • Diastolic blood pressure test of 90 mmHg or above • Treatment for hypertension
'At risk of' hypercholesterolaemia	Patients with one or more codes on any of the following lists: • At risk of hypercholesterolaemia • Total cholesterol level test between 4-5 mmol/L • LDL level test between 2-3 mmol/L
Hypercholesterolaemia	Patients with one or more codes on any of the following lists: • Hypercholesterolaemia • Total cholesterol level test above 5 mmol/L • LDL level test above 3 mmol/L • Treatment for hypercholesterolaemia
'At risk of' type 2 diabetes	Patients with one or more codes on the list 'at risk of type 2 diabetes' without codes on the list 'type 1 diabetes'.
Type 2 diabetes	Patients with one or more codes on any of the following lists: • At risk of type 2 diabetes • Type 2 diabetes • Treatment for type 2 diabetes • Unspecified diabetes, if recorded after age 40 • Treatment with insulin, if recorded after age 40 Without codes on the list 'type 1 diabetes'.
Dementia	Patients with one or more codes on any of the following lists: • Possible AD • Probable AD • Non-specific dementia • Other dementia • Vascular dementia • Treatment for dementia
Parkinson's disease	Patients with one or more codes on any of the following lists: • Parkinson's disease • Treatment for Parkinson's disease
Amyotrophic lateral sclerosis	Patients with one or more codes on any of the following lists: • Amyotrophic lateral sclerosis • Treatment for amyotrophic lateral sclerosis

Figure 5.1: Decision tree for determining dementia subtype for analyses using the CPRD in this thesis.



Note: some analyses refer to the category 'Non-AD and mixed dementias' or 'Non-AD dementia'. These categories combine individuals from 'Vascular dementia' and 'Other dementias' in a single group. Patients in these groups may also have 'Alzheimer's disease' codes recorded. Analyses referring to 'any dementia' are using the general definition of dementia given in Table 5.1, which combines all dementia subtypes.

indicated a diagnosis by restricting the events to the 'enttype' and data combinations of interest. The result of this is an event list equivalent to those created using the codes lists that are described above. As before, I created a patient list from the event list that contained each patient only once with the date of their first recorded code and the number of entries they have on the event list.

Once all the event and patient lists were created, I could refer to them when constructing the cohorts for each of the analyses detailed in this thesis. Analysis-specific use of these code lists is detailed in the relevant chapters.

5.8. Defining diagnoses

The diagnoses in this thesis are defined in Table 5.1. It is assumed that treatment implies diagnosis, so a diagnosis may be defined by both diagnostic and treatment events. All of the code lists referred to in the table are available from the referenced online source. (150) For some of the analysis presented in this thesis, it was necessary to define dementia by its subtypes. For example, when studying the currently licensed treatments as in Chapter 6, which are only indicated for use by those with Alzheimer's disease. The decision tree used to define dementia subtype is presented in Figure 5.1.

5.9. ICD-10 equivalences

As explained above, diagnosis definitions in the CPRD are determined by Read codes, whereas both the ONS death registry and the HES inpatient dataset use codes from the International classification of diseases (ICD). The ICD is maintained by the World Health Organization and is currently in its 10th revision: ICD-10. To use this data, I created ICD-10 code lists that correspond to the Read code lists used for the CPRD data extract. As before, these code lists were then refined under the guidance of my supervisors and have been made available online. (149)

ICD-10 and Read codes do not map to each other exactly, with ICD-10 codes generally covering multiple Read codes. As I had been conservative and specific with the approach to the Read codes, I included ICD-10 codes on multiple code lists where appropriate. This helped to ensure the scope of the Read code lists was covered when using the less

specific ICD-10 codes. For example, the ICD-10 code 'F03' represents 'Unspecified dementia' and is defined as including the following diagnoses, many of which are ‘not otherwise specified’ (NOS):

- Presenile dementia NOS
- Presenile psychosis NOS
- Senile dementia NOS
- Primary degenerative dementia NOS
- Senile dementia, depressed or paranoid type
- Senile psychosis NOS

There are Read codes for each of the above bullet points and for 'Unspecified dementia'. In the Read code lists, 'Unspecified dementia' and 'Primary degenerative dementia NOS' were assigned to the non-specific dementia code list and the remaining codes to the possible Alzheimer’s disease code list. I therefore chose to include the ICD-10 code ‘F03’ for ‘Unspecified dementia’ on both the non-specific dementia and possible Alzheimer’s disease ICD-10 code lists to account for the multiple Read codes it relates to. Equivalent code lists were required because both the ONS and HES databases are used for sensitivity analyses to assess the sensitivity and specificity of diagnoses definitions in this thesis (Section 6.3.4.2).

5.10. Covariates

Some analyses presented in this thesis adjust for covariates, all of which are listed in Table 5.2. Collider bias (Section 2.8.1.2) could occur if events that happened because of the prescription the patient was issued are conditioned on. To prevent this bias from affecting the results, covariates are defined using data inputted prior to the first prescription. (66) All of the code lists referred to in the table are publicly available: http://rebrand.ly/repurposing-antihypertensives-dementia-codelists.

5.11. Summary

This chapter introduced the CPRD, an ongoing UK-based primary care database containing over 11.3 million patients, which provided the electronic health record data for this thesis. It included information on the data structure; rationale for the linkages to the ONS death registry, the HES database and the Index of Multiple Deprivation; details

on how the data were cleaned; and definitions for the diagnoses, treatments, and covariates used in this thesis. The data described in this chapter will be used in Chapter 6 to examine prescribing trends for dementia drugs and Chapter 7 to assess whether antihypertensives can be repurposed for dementia prevention.

Table 5.2: Covariate definitions for analyses using the CPRD in this thesis.

Covariate	Definition
Body mass index	Defined as the value recorded in the additional clinical details files or calculated using the most recent height and weight measurements. Measurements are restricted to those taken over age 25 to ensure they are adult measurements.
Smoking status	Defined as a three-factor variable: current, former or never smoker. Variable assigned using the additional clinical details files and code lists obtained from the referenced study. (28) If conflicting records of smoking status exist, current smoker takes precedence over other statuses, followed by former smoker.
Alcohol consumption	Defined using the additional clinical details files, which provides a measure of alcohol consumption in units per week.
Consultation rate	Calculated by dividing the total number of clinic visits by the length of time in the dataset prior to a patient's index date. This gives an estimate of average consultations per year.
Socioeconomic position	Determined using linked data from the Index of Multiple Deprivation, where 1 = least deprived and 20 = most deprived.
Previous history of coronary-artery disease	Defined using a Read code list, refined under the guidance of a practicing GP.
Previous coronary-bypass surgery	Defined using a Read code list, refined under the guidance of a practicing GP.
Cerebrovascular disease including stroke	Defined using a Read code list, refined under the guidance of a practicing GP.
Other major chronic illness	Defined using an adapted Charlson index that included the following conditions: rheumatological disease, peptic ulcer disease, myocardial infarction, congestive heart disease, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, mild liver disease, diabetes, hemiplegia, renal disease, diabetes with complications, cancer, moderate liver disease, metastatic tumour, and AIDS. (141,142) This covariate is included in the analysis because people who have more chronic conditions, such as diabetes, may have higher rates of consultation, which may also increase the opportunity for recording other diagnoses such as dementia. This is known as ascertainment bias (Section 2.8.1.1).

Chapter 6. Results: Prescribing trends for drugs for dementia

6.1. Introduction

There are currently four licensed treatments that provide symptomatic relief for patients with Alzheimer's disease in England - three acetylcholinesterase inhibitors (donepezil, rivastigmine, galantamine) and one NMDA receptor antagonist (memantine). These drugs are collectively referred to as 'drugs for dementia' in the British National Formulary, despite their licensing for Alzheimer's disease only. (160) Since the first of these drugs became available in 1997, there have been several changes in national guidelines for the treatment of Alzheimer's disease, as well as several initiatives to encourage better diagnosis and treatment of the disease. Despite this, there has been little research into whether such changes to guidelines and initiatives have directly influenced clinical practice. (33,34) This chapter examines how prescription rates in England have changed in response to these factors since their launch up to 1st January 2016. As detailed in Section 1.6, the motivation for this chapter was primarily to identify factors that may influence prescribing in this therapy area and consequently effect repurposed drug candidates in the future. However, the analysis also provides other useful information such as when breaks in the prescription of these drugs have occurred, which could be exploited as natural experiments for progression studies. The analyses presented in this chapter, conducted under the supervision of Neil Davies, Richard Martin and Patrick Kehoe, have been published in Alzheimer's Research & Therapy. (35)

6.2. Background

This chapter will focus on factors at a national level that may have influenced prescribing. Specifically, it will look at how prescribing was affected by changes in NICE guidance (including the 2006 guidance that was subject to legal challenges); the addition of dementia to the QOF; the introduction of ambitious government dementia strategies and the expiry of drug patents. The timings of each of these changes, which may have influenced aspects of drug prescribing and clinical practice, are discussed further below and summarized in Table 6.1. To my knowledge, there were two existing studies that considered prescribing trends at the time I conducted this study and they mainly focused on the impact of the National Dementia Strategy. (33,34) This study extends the findings

of the previous studies because it considers trends since the launch of these drugs and allows consideration of multiple factors affecting prescribing by using a joinpoint model (Section 6.3.3) as a hypothesis-free approach.

Table 6.1: Events potentially influencing dementia drug prescribing in England that occurred between May 1997 and January 2016.

Event Date	Event
May 1997	Donepezil first recorded in CPRD
September 1998	Rivastigmine first recorded in CPRD
January 2001	Galantamine first recorded in CPRD First NICE guidance released
December 2002	Memantine first recorded in CPRD
November 2006	NICE recommend restricting drug access
September 2007	QOF revised to include dementia
February 2009	First National Dementia Strategy launched
March 2011	NICE remove recommendation restricting drug access
January 2012	Galantamine patent expired
February 2012	Donepezil patent expired
May 2012	Prime Minister's Dementia Challenge launched
July 2012	Rivastigmine patent expired
April 2014	Memantine patent expired
February 2015	Prime Minister's Challenge on Dementia 2020 launched

6.2.1. NICE guidance on the prescribing of drugs for dementia

In the past NICE guidance has used scores from the mini mental state examination (MMSE), in combination with other measures, to guide whether a patient should be prescribed a drug for dementia. The test, proposed in 1975 by Folstein et al, scores a patient's cognition out of 30, where normal cognition is considered as a score of 24 or more. (161) The original NICE guidance, issued in 2001, on the use of drugs to treat Alzheimer's disease recommended that the three acetylcholinesterase inhibitors should be used for all patients scoring 12 or above on the MMSE until the drugs were deemed no longer effective. (162,163) In November 2006, NICE revised their guidance so that the use of acetylcholinesterase inhibitors was restricted to patients with moderate Alzheimer's disease – this was defined as patients scoring between 10 and 20 points on the MMSE. The 2006 guidance was also the first to consider the use of the NMDA receptor antagonist, memantine, which was recommended for use only in clinical trials for patients with moderate to severe disease. (164) This revision of the guidance was

controversial due to the way in which it assessed cost effectiveness, which was expected to restrict access to these drugs, and was ultimately the subject of a high court challenge by the Alzheimer's society and two drug manufacturers: Eisai and Pfizer. (165–167) This led to a further revision being made to the NICE guidance at the end of March 2011, which recommended acetylcholinesterase inhibitors for patients with mild to moderate Alzheimer's disease and memantine for patients with moderate to severe Alzheimer's disease, or who could not tolerate acetylcholinesterase inhibitors. (168) For the duration of this study, treatment had to be initiated by a specialist and was deemed effective as long as there had been "an improvement or no deterioration in MMSE score, together with evidence of global improvement on the basis of behavioural and/or functional assessment." (163)

6.2.2. Inclusion of dementia on the QOF

QOF is a voluntary incentive program, introduced in 2004, to improve services in primary care. (169) Dementia first appeared in QOF as an 'indicator' in September 2007. (170) There are currently three indicators for dementia included in the framework. The first requires that the practice establish and maintain a register of patients diagnosed with dementia and the further two indicators refer to the ongoing management of the disease. (171) The inclusion of dementia on the QOF could have encouraged more of a focus on the diagnosis and pharmacological management of the disease in participating practices.

6.2.3. Government dementia strategies

The first National Dementia Strategy was launched by the Department of Health in February 2009. The aim of that strategy was "to ensure that significant improvements are made to dementia services across three key areas: improved awareness, earlier diagnosis and intervention, and a higher quality of care". (172) This strategy was followed in 2012 by the Prime Minister's Dementia Challenge, which looked to improve care and research by 2015, and more recently the Prime Minister's Challenge on Dementia 2020. (173,174) The most recent strategy aims to build on the work of its predecessors to make England the best place for both dementia care and research. In general, such strategies may help to increase the awareness of dementia for both the public and health services. (175,176)

6.2.4. Drug patents

The charity King's fund found that the prescription of generic drugs over their patented alternatives has "saved the NHS [National Health Service] around £7.1 billion and allowed more than 490 million more items to be prescribed to patients" between 1976 and 2013. (177) Acetylcholinesterase inhibitors for the treatment of Alzheimer's disease became available generically from 2012, while NMDA receptor antagonists became available generically from 2014. (178) Therefore, in recent years the cost of drugs for dementia has reduced significantly from previous years. This serves as a potential factor in rates of prescribing, particularly in publicly funded health care services such as the National Health Service in England. The patent information for each of the individual drugs is provided in Table 6.2.

Table 6.2: Patent information for the British National Formulary category 'drugs for dementia'.

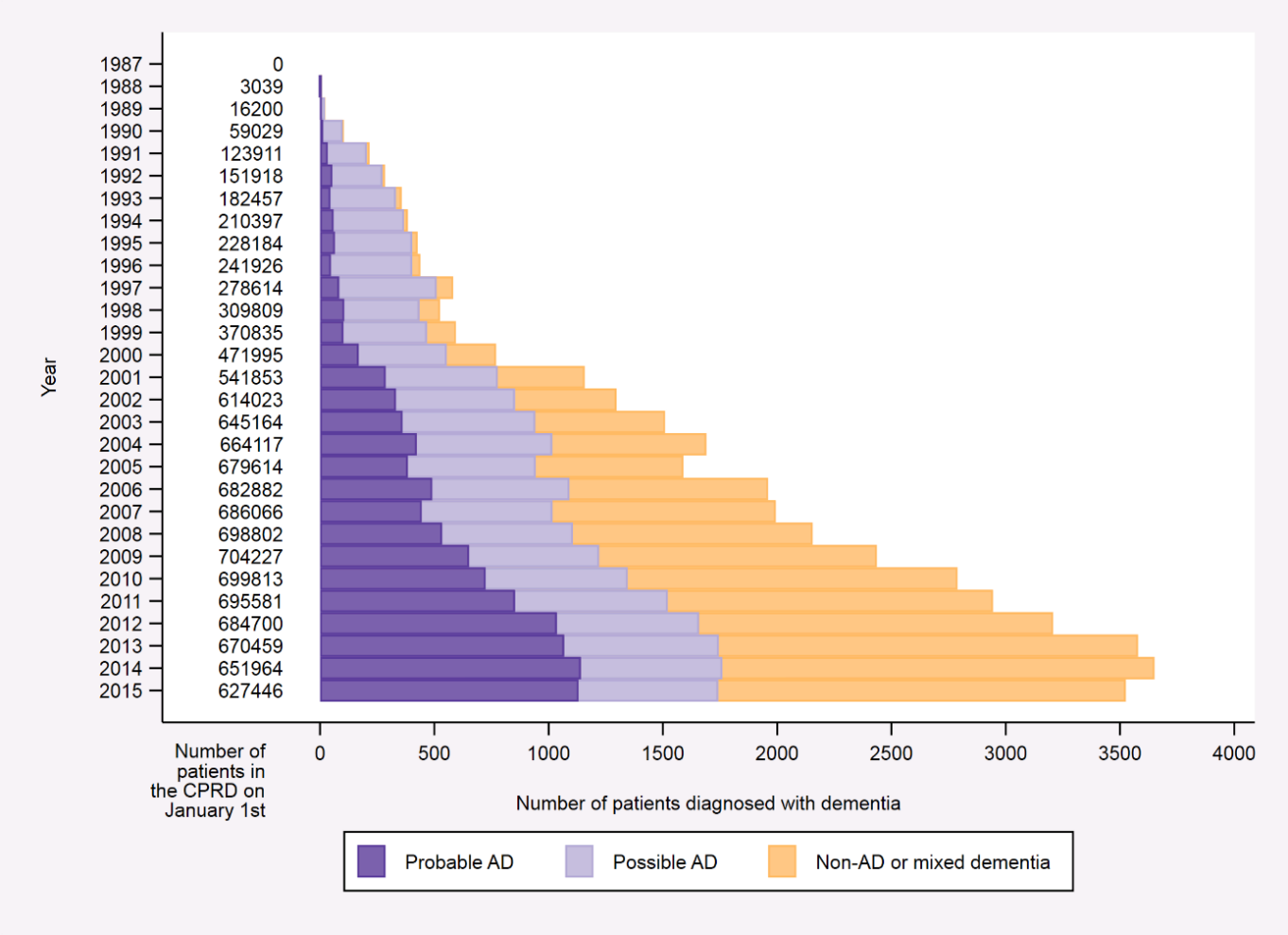
Generic name	Patent name	Drug class	Patent expiry
Donepezil	Aricept	Acetylcholinesterase inhibitor	January 2012
Rivastigmine	Exelon	Acetylcholinesterase inhibitor	February 2012
Galantamine	Reminyl	Acetylcholinesterase inhibitor	July 2012
Memantine	Ebixa	NMDA receptor antagonist	April 2014

6.3. Methods

6.3.1. Study design and setting

This study was a joinpoint analysis of the proportion of patients, who were eligible for treatment and had a diagnosis of 'probable Alzheimer's disease', that received their first prescription for the treatment of interest in each month in the CPRD. The CPRD is described fully in Chapter 5. Patients were defined as eligible for their first prescription if they had the diagnosis of interest with no previous prescription for the treatment of interest. The time period was measured in units of one month as this was the smallest clinically meaningful measure that could realistically be defined. I investigated treatment rates as a proportion of eligible patients, because the underlying rate of diagnosis of Alzheimer's disease, as well as non-Alzheimer's disease and mixed dementias, has changed over time in the CPRD. This is illustrated in Figure 6.1 using patients who received a diagnosis prior to 1st January 2016 and are from an English practice with a last data collection date in 2016 to reflect the present study.

Figure 6.1: Patients diagnosed with dementia in English practices within the CPRD between 1987 and 2015.



The data presented are restricted to patients who received a diagnosis prior to 1st January 2016 and are from an English practice with a last data collection date in 2016 to reflect the present study.

The four drugs for dementia were separated according to drug class, i.e. acetylcholinesterase inhibitors and NMDA receptor antagonists for the analysis. Exposure date was taken to be the date on which the first prescription requesting the drug(s) being considered was recorded. This allowed patients who had previously been prescribed acetylcholinesterase inhibitors to be included in the NMDA receptor antagonist analysis. This is necessary as NMDA receptor antagonists may be prescribed alongside acetylcholinesterase inhibitors and are often given to patients later in the course of their disease, potentially following exposure to acetylcholinesterase inhibitors.

6.3.2. Participants

Patients were included in this study if they had a diagnosis of dementia as defined in Section 5.8; had data available between January 1st 1987 and December 31st 2015; and attended an English practice with a last data collection date in 2016. The latter two conditions ensured all data were complete for the timeframe being considered and were subject to the guidelines and initiatives being studied. Diagnosis date was taken to be the first date on which a code relating to dementia, regardless of whether it was for a diagnosis or a treatment, was recorded.

6.3.3. Statistical analysis

As highlighted earlier, this study used joinpoint analysis. A joinpoint can be thought of as a break point between two fitted models, which allows the models on either side to take different values. Joinpoint analysis works by adding one joinpoint at a time and testing for statistical significance using the Monte Carlo Permutation method. Once adding joinpoints no longer effects the result, or the maximum number of pre-specified joinpoints is reached, the analysis is complete. (179) It is usual practice to start with zero joinpoints – that is to have one model that is fitted to all the data. Joinpoint analysis has been developed for incidence rates and so prevalent drug use, which requires consideration of both incidence and continued drug use, cannot be studied in this way.

The analysis of each treatment of interest in this study started on the first day of the month following the first recorded prescription for that treatment. For example, the first prescription for NMDA receptor antagonists occurred on 16th December 2002 and so the

analysis of this drug class started on 1st January 2003. For each patient, the month and year of both their diagnosis and their first prescription were used. For each month, I calculated: (A) the number of patients receiving their first prescription in that month; and (B) the number of patients with a diagnosis who had not received treatment before the first of the month. Dividing A by B provided the proportion of patients with diagnoses who received their first prescription for the treatment of interest each month. I also calculated the standard error of this proportion according to the following formula: (180)

$$SE_i = \sqrt{\frac{p_i(1-p_i)}{n_i}}$$ for proportion p and sample size n for observation i

The trend analysis using joinpoint models was then conducted. The optimum number of joinpoints, as determined by the software and up to a maximum number of two, was used to select the model. The period between two joinpoints is referred to as a 'segment' and they are numbered chronologically. The model used assumes that the rate of prescription "changes at a constant percentage of the rate of the previous year" and so is determined by the following equation: $\ln y = xb$. This allows consideration of the monthly percent change. The trend over the entire study period is summarized using the average monthly percent change. This is calculated as the average of the monthly percent changes, weighted by segment length. (179) All analysis was conducted using Joinpoint Regression Program (version 4.3.1.0) and Stata (version 14.1). (80,181) The model specifications for the joinpoint analyses are the software's default with dependent variable type set to 'proportion' and the maximum number of join points set to two. The Stata code used in this analysis is available from GitHub: https://github.com/venexia/DementiaDrugsCPRD.

6.3.4. Sensitivity analyses

6.3.4.1. Representativeness of the study population compared with the CPRD as a whole

As highlighted in Section 5.2, the CPRD is "broadly representative of the UK general population". (147,153) However, I was concerned that restricting the study population to those from English practices might have jeopardised its representativeness. I therefore

compared the distribution of patients and practices eligible for the study by nation with those eligible for the study in the complete dataset to see whether the distributions were maintained. I also compared the distribution of patients and practices eligible for the study between regions in England to see if any of the other selection criteria, such as a last data collection date in 2016, could have influenced the results.

6.3.4.2. Sensitivity and specificity of the diagnosis definitions

It is difficult to differentiate dementia subtypes however, drugs for dementia are currently only licensed for the treatment of Alzheimer's disease. Consequently, it was important to know the sensitivity and specificity of the diagnosis definitions – particularly if I wanted to identify any off-license use of these drugs for other dementia subtypes. Sensitivity indicates whether patients in the CPRD have been 'correctly diagnosed' based on the information in the comparator dataset, while specificity indicates whether patients have been 'correctly not diagnosed'. I was able to calculate these quantities using linked data from the ONS death registry and the HES inpatient dataset as the comparators for the data from the CPRD. HES outpatient data was excluded from the sensitivity analysis as it is known that less than 5.0% of patients have a diagnosis recorded in this dataset. (182) Note that I do not believe the ONS death registry and the HES inpatient dataset to represent a 'gold standard' for this comparison as both will be subject to missing data. However, these datasets do offer a secondary source of data to that from the CPRD, which consists of the same patients but is likely to be subject to different recording issues.

First, I defined the following four terms based on where a patient had codes recorded:

- True positive: the code is in the CPRD and is in the linked data
- False positive: the code is not in the CPRD, but is in the linked data
- True negative: the code is not in the CPRD and is not in the linked data
- False negative: the code is in the CPRD, but is not in the linked data

I then calculated the sensitivity and specificity (183):

$$\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$\text{Specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$

6.3.4.3. Alternative outcomes

I repeated the main analysis, which considers the diagnosis 'probable Alzheimer's disease', with relaxed diagnosis definitions to test the sensitivity of the results to this outcome definition. I did this in two ways: (1) introducing codes that represented what may be lesser degrees of confidence in the accuracy of Alzheimer's disease diagnosis (termed 'any Alzheimer's disease') and (2) introducing codes capturing other types of dementia (termed 'any dementia'). By repeating the same analysis with different definitions, I was able to assess whether the joinpoints selected in the model were consistent across the definitions.

6.3.5. News search

Several of the national guidelines and initiatives considered in this study may have increased awareness of dementia, including Alzheimer's disease. To investigate this, I downloaded the Google trends data for news searches in England for the disease term 'Alzheimer's Disease' from 1st January 2008 up to 1st January 2016. (184) Unfortunately, data were not recorded prior to 2008 and so I cannot comment on the effect media coverage may have had on trend changes identified before this point in time.

6.4. Results

6.4.1. Trend analysis for acetylcholinesterase inhibitors

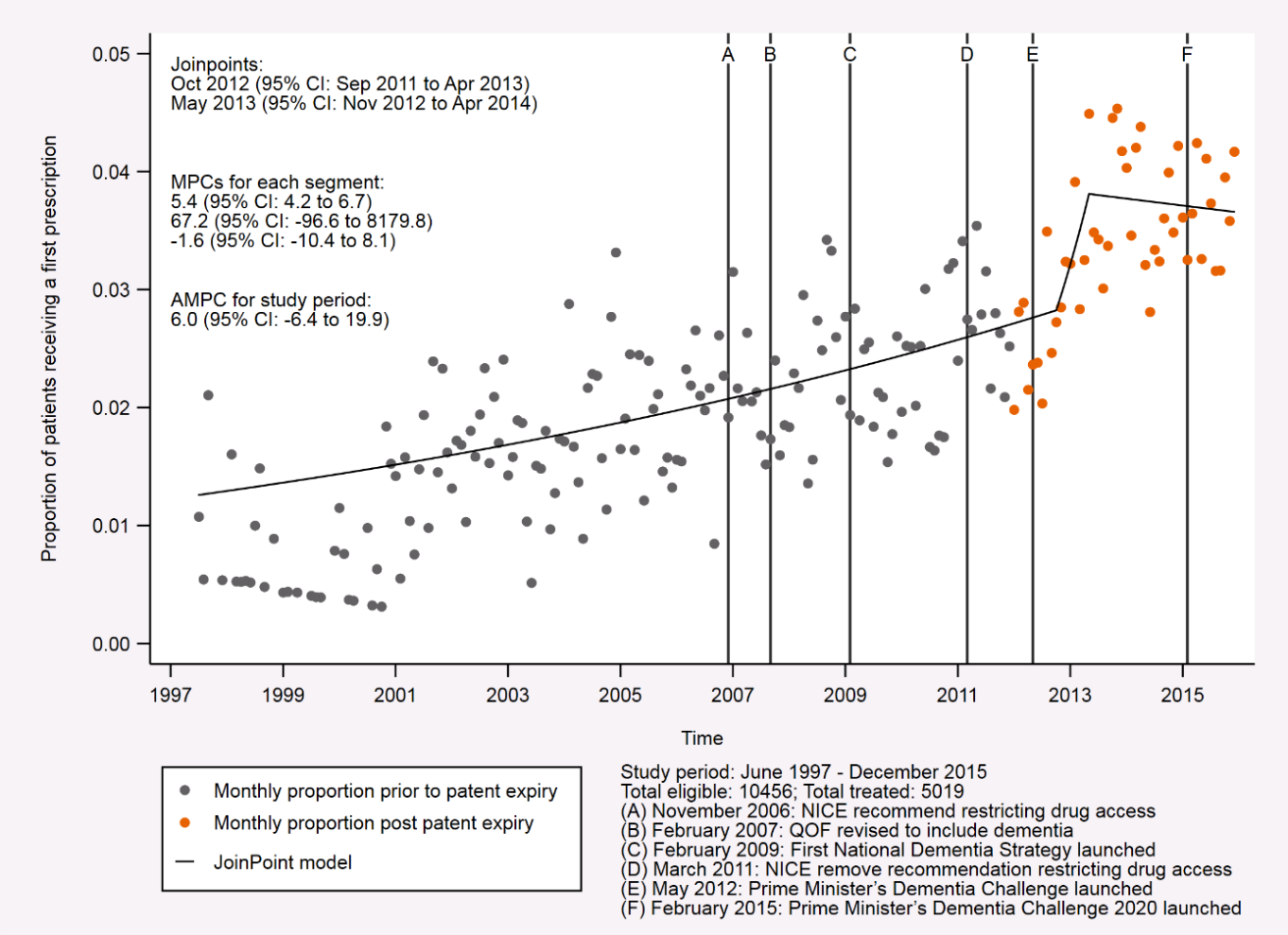
The average monthly percent change for the proportion of patients with probable Alzheimer's disease receiving their first prescription for an acetylcholinesterase inhibitor during the study period was 6.0 (95% CI: -6.4 to 19.9) (Figure 6.2). This represents the weighted average for three trends that occurred between June 1997 and December 2015. The first trend was for a steady increase in the proportion of patients to receive their first prescription for an acetylcholinesterase inhibitor. This is reflected in the monthly percent change of 5.4 (95% CI: 4.2 to 6.7). The second trend was for a surge in the prescription

rate, which corresponded to a monthly percent change of 67.2 (95% CI: -96.6 to 8179.8) – note that this estimate is very uncertain because it is only based on seven data points. The final trend had a monthly percent change of -1.6 (95% CI: -10.4 to 8.1), falling below zero for the first time since the launch of these drugs. The trend changes were estimated to have occurred in October 2012 (95% CI: September 2011 to April 2013), and, less than a year later, in May 2013 (95% CI: November 2012 to April 2014).

6.4.2. Trend analysis for NMDA receptor antagonists

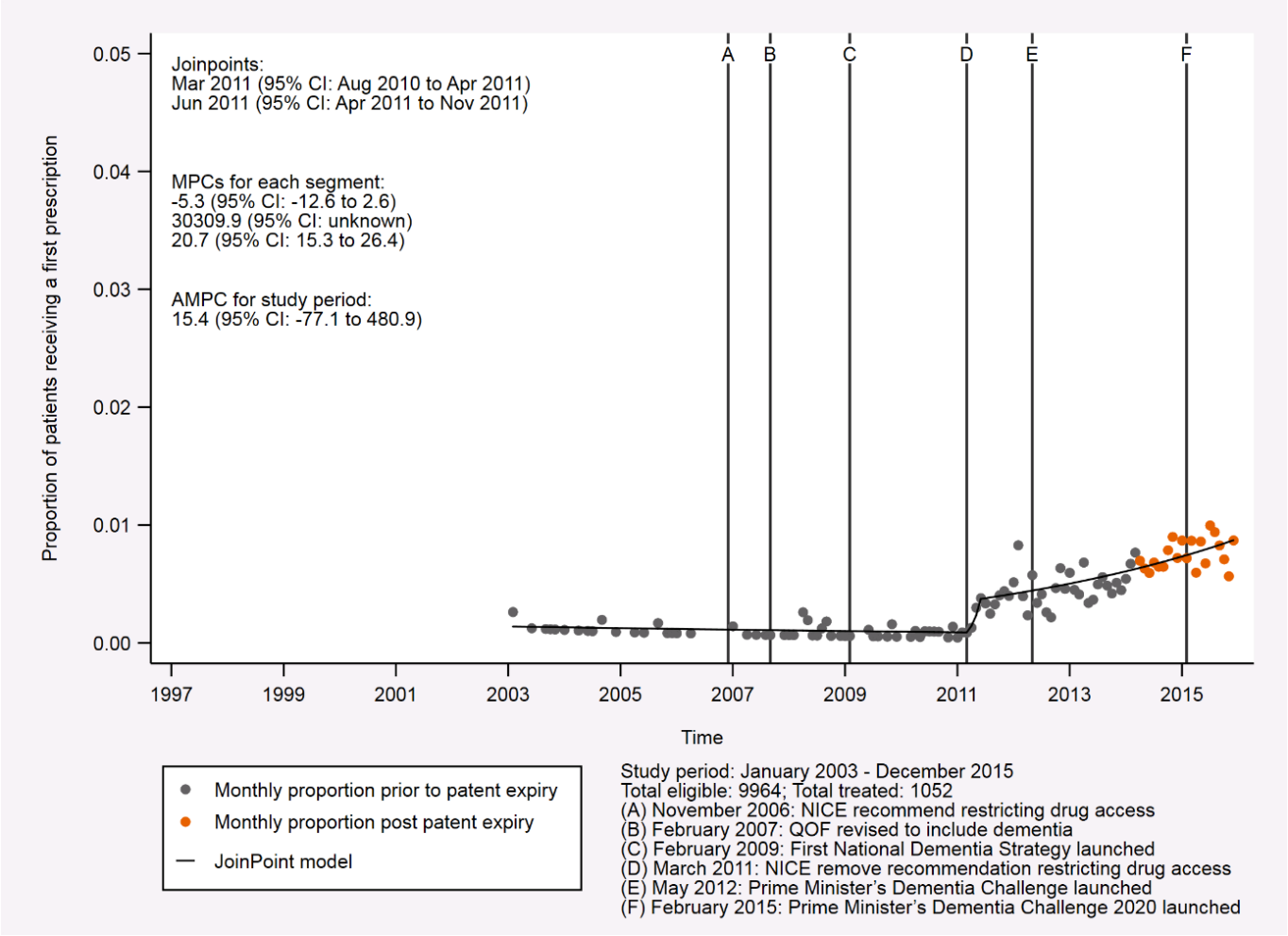
Figure 6.3 presents the equivalent analysis for the NMDA receptor antagonist, memantine. Memantine became available in January 2003 and was prescribed much less than the other drugs, despite similar numbers of eligible patients. This is partly related to the indication of these drugs. Memantine was originally recommended for more advanced disease than the acetylcholinesterase inhibitors and is often added to a prescription of acetylcholinesterase inhibitors following progression of the disease. Despite this, as observed for the acetylcholinesterase inhibitors, the point estimate for the average monthly percent change in the proportion of patients with probable Alzheimer's disease receiving their first prescription for an NMDA receptor antagonist was positive (15.4; 95% CI: -77.1 to 480.9), though the 95% CI around this estimate was large. The initial trend for prescribing of this drug showed a reduced number of prescriptions in the time that followed the launch with a monthly percent change of -5.3 (95% CI: -12.6 to 2.6). This changed around March 2011 (95% CI: August 2010 to April 2011) to an increasing trend with a monthly percent change of 30309.9 (95% CI: unknown). This estimate for the monthly percentage change is unreliable as it is based on just three data points – specifically the monthly estimates for March 2011, April 2011, and May 2011. However, it indicates a distinct change from the previously decreasing trend for prescribing these drugs. Following this, in June 2011 (95% CI: April 2011 to November 2011) until the end of the study in December 2015, this trend reduced to a monthly percent change of 20.7 (95% CI: 15.3 to 26.4) .This indicates a continuing increase in the prescriptions for NMDA receptor antagonists in recent years, albeit substantially reduced from the rise observed between March and June 2011.

Figure 6.2: Acetylcholinesterase inhibitor prescriptions in patients with probable Alzheimer's disease in the CPRD between June 1997 and December 2015.



This graph shows the proportion of patients with probable Alzheimer's disease receiving their first prescription for an acetylcholinesterase inhibitor each month from June 1997 to December 2015. The fixed lines indicate events with the potential to effect prescription rates during the study period. The joinpoints, monthly percent change (MPC) for each segment and the average monthly percent change (AMPC) for the entire study period are also presented.

Figure 6.3: NMDA receptor antagonist prescriptions in patients with probable Alzheimer's disease in the CPRD between January 2003 and December 2015.



This graph shows the proportion of patients with probable Alzheimer's disease receiving their first prescription for an NMDA receptor antagonist each month from January 2003 to December 2015. The fixed lines indicate events with the potential to effect prescription rates during the study period. The joinpoints, monthly percent change (MPC) for each segment and the average monthly percent change (AMPC) for the study period are also presented.

6.4.3. Sensitivity analyses

6.4.3.1. Representativeness of the study population compared with the CPRD as a whole

There is a total of 135,144 patients from 675 practices that have a diagnosis of dementia recorded in the CPRD according to the definition in Section 5.8. Table 6.3 and Table 6.4 show the distribution of these patients and practices, respectively, at the national level. Out of the 135,144 patients recorded as having dementia in the CPRD, 47.2% of patients qualify for the study. In England, this percentage is slightly lower at 38.8% of patients. Given the similarity of these percentages and the fact that 76.7% of patients in the CPRD are registered at an English practice, the English population is likely to be representative of the CPRD dataset as a whole. Table 6.4 confirms the observations from Table 6.3 with the distribution of practices mostly in line with the distribution of patients, providing further evidence for the English practices being representative of the CPRD as a whole. It is worth noting that the percentage of patients qualifying for the study in other nations of the UK is much higher than in England. This is due to the other criteria, namely that practices should have their last data collected in 2016. While many practices outside of England met this criterion, data collection is less consistent in England. This is discussed in further detail below.

Table 6.5 and Table 6.6 show the distribution of patients and practices within England respectively at the regional level. Ideally, there would be consistency in both the percentage of patients and practices that meet the study criteria, as this would preserve the representativeness of the CPRD dataset. The South East Coast has both the most practices and patients that meet the study criteria. On the other hand, the East Midlands does not have any patients that qualify for the study. This is because, in the data extract (March 2016 snapshot), there has been no data collected from a practice in the East Midlands since 2014. Despite these regions, the average percentage of patients included in the study is 38.8% (interquartile range: 19.2 to 41.0). Similarly, the percentage of practices included in the study is 37.4% (interquartile range: 19.5 to 45.2). This indicates reasonable consistency across the dataset. Furthermore, the consistency must be weighed against other factors, such as data quality. For this reason, the slight compromise in consistency resulting from the exclusion of practices in the East Midlands and elsewhere, due to a lack of recent data, is passable.

Table 6.3: The distribution of patients by nation in the prescribing trends study.

Nation	Study	CPRD	Percentage
England	40202	103595	38.8
Northern Ireland	4380	5146	85.1
Scotland	11114	15214	73.1
Wales	8090	11159	72.5
Total	63786	135114	47.2

Table 6.4: The distribution of practices by nation in the prescribing trends study.

Nation	Study	CPRD	Percentage
England	195	522	37.4
Northern Ireland	21	22	95.5
Scotland	67	79	84.8
Wales	44	52	84.6
Total	327	675	48.4

Table 6.5: The distribution of patients by region in England in the prescribing trends study.

Region	Study	CPRD	Percentage
North East	504	2365	21.3
North West	6383	15914	40.1
Yorkshire and the Humber	561	5109	11.0
East Midlands	0	4574	0.0
West Midlands	3731	11552	32.3
East of England	1975	10435	18.9
South West	4718	12587	37.5
South Central	7659	16041	47.7
London	5544	11817	46.9
South East Coast	9127	13201	69.1
Total	40202	103595	38.8

Table 6.6: The distribution of practices by region in England in the prescribing trends study.

Region	Study	CPRD	Percentage
North East	3	12	25.0
North West	29	80	36.3
Yorkshire and the Humber	3	29	10.3
East Midlands	0	25	0.0
West Midlands	20	58	34.5
East of England	9	52	17.3
South West	19	60	31.7
South Central	28	54	51.9
London	37	87	42.5
South East Coast	47	65	72.3
Total	195	522	37.4

6.4.3.2. Sensitivity and specificity of the diagnosis definitions

As stated in Section 6.3.4.2, I do not believe the ONS death registry and the HES inpatient dataset to represent a 'gold standard' against which I can calculate the sensitivity and specificity of the diagnosis definitions. However, these data sources do allow comparison of the diagnoses recorded for the same set of patients and are therefore a useful validation of the CPRD results. To be included in this analysis, patients had to have linkage across the three datasets, namely: CPRD Gold, the HES inpatient dataset, and the ONS death registry. Consequently, the 29,362 patients included in this analysis are all deceased and so the recorded diagnoses can be considered final.

Table 6.7 presents the sensitivity and specificity of the diagnoses using the HES inpatient dataset as the comparator. Sensitivity for the diagnosis possible Alzheimer's disease is poor (37.3%), however the other diagnoses perform better with non-AD and mixed dementias performing the best (71.6%). Sensitivity estimates the proportion of people in the CPRD that have been 'correctly diagnosed' based on the information in the HES inpatient dataset. Therefore, the estimate of 37.3% indicates that 37.3% of people identified as having possible Alzheimer's disease in the CPRD had the same diagnosis recorded in the HES inpatient dataset. The specificity of the diagnoses is generally much better across all diagnoses (≥ 62.9%) with the diagnosis probable Alzheimer's disease being the most specific (79.1%). Specificity estimates the proportion of people in the CPRD that have been 'correctly not diagnosed' based on the information in the HES inpatient dataset. Therefore, the estimate of 79.1% indicates that 79.1% of people identified as not having probable Alzheimer's disease in the CPRD were also recorded as not having probable Alzheimer's disease in the HES inpatient dataset. The higher specificity of the code lists reflects the conservative approach taken with the Read code lists, which are used in combination to determine diagnosis. While sensitivity is low for the Alzheimer's disease diagnoses, particularly the possible cases, the larger sample size used in the study (that includes patients without linked data) potentially provides greater power, even if some patients are missed.

Table 6.8 presents the sensitivity and specificity of the diagnoses using the ONS inpatient dataset as the comparator. The general pattern across diagnoses is much the same as that observed in the analysis that used the HES inpatient dataset as the comparator. Both

probable Alzheimer's disease and non-AD and mixed dementias have slightly higher sensitivity but lower specificity than the previous analysis. Meanwhile possible Alzheimer's disease performs slightly worse for both sensitivity and specificity. This is likely due to the ONS death registry recording diagnoses at the time of, or after, death. Because of this, you might expect possible diagnoses to be less relevant and potentially other forms of evidence to be available to preclude conditions (for example, through post mortem). Overall, both these comparisons indicate variable sensitivity with high specificity of the diagnoses. Encouragingly, the definition of probable Alzheimer's disease, which is used in the main analysis, performs well against both comparators.

6.4.3.3. Alternative outcomes

The results of the analysis with the alternative, more relaxed, outcome definitions can be found in Table 6.9. The table shows that the joinpoint analysis is consistent regardless of the diagnosis definition used for NMDA receptor antagonists. On the other hand, the joinpoint estimates for the acetylcholinesterase inhibitors vary according to the diagnosis definition used, though the two sensitivity analyses are reasonably consistent with each other.

6.4.4. News search

Figure 6.4 presents the Google trends data for news searches in England for the disease term 'Alzheimer's Disease' each month from January 2008 to December 2015 inclusive. There were no strong trends in the interest for the search term with values indicating both low and high interest occurring throughout the period studied. Months with insufficient data, indicating little interest in the search term, became less common over the period studied – the most recent occurring in August 2015. Interest peaked in September 2012 and was also high in January 2011 (88%), January 2010 (82%) and April 2008 (81%).

Table 6.7: Sensitivity and specificity of diagnoses in the CPRD compared against HES.

	Patients in CPRD dataset	Patients in HES inpatient dataset	Sensitivity	Specificity
Probable AD	8069	5007	59.4	79.1
Possible AD	8259	6461	37.3	74.5
Non-AD and mixed dementias	13034	6206	71.6	62.9

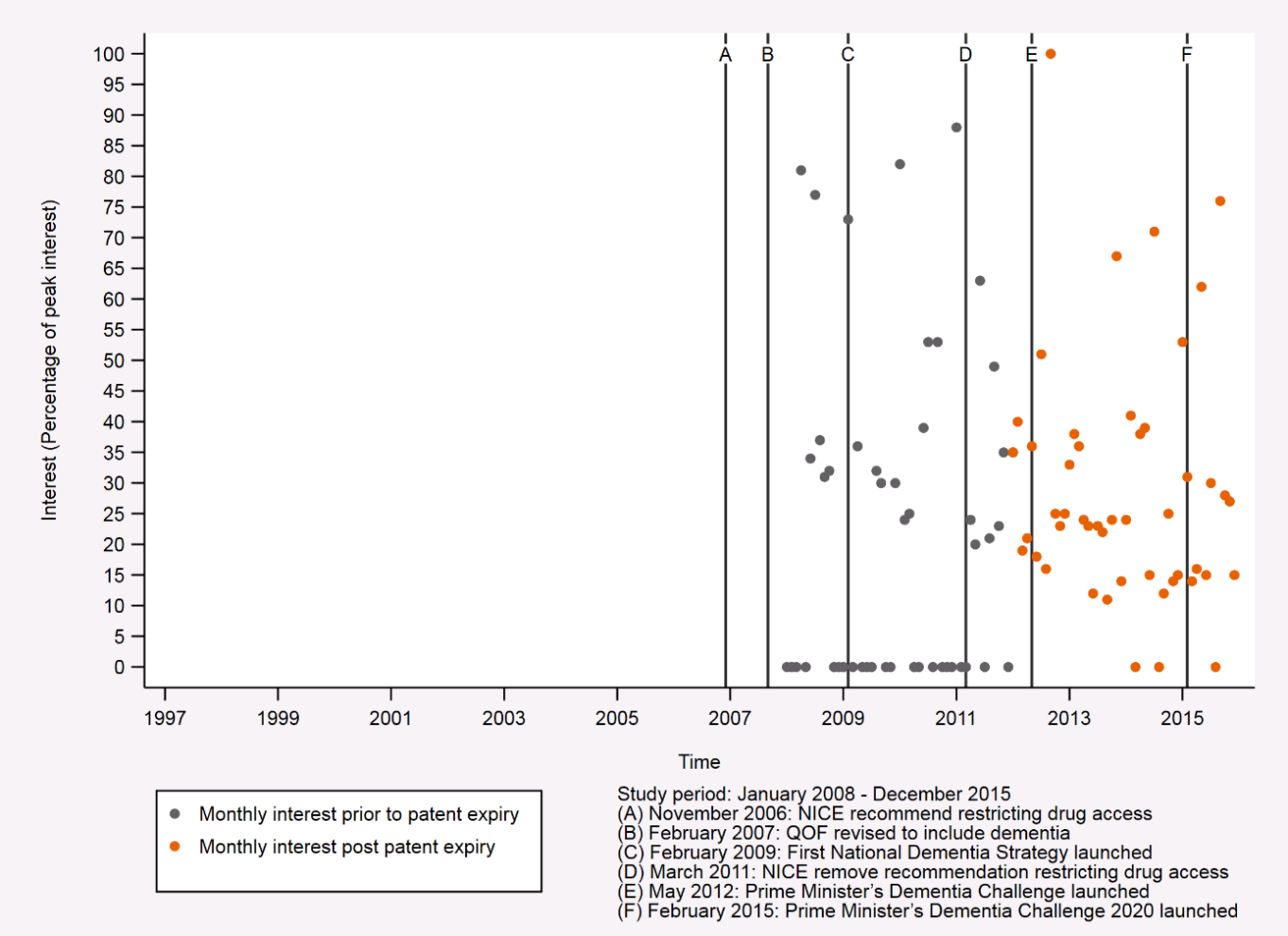
Table 6.8: Sensitivity and specificity of diagnoses in the CPRD compared against ONS.

	Patients in CPRD dataset	Patients in ONS death registry	Sensitivity	Specificity
Probable AD	8069	1863	65.5	75.1
Possible AD	8259	4752	36.0	73.4
Non-AD and mixed dementias	13034	1456	80.4	57.5

Table 6.9: Comparison of trend analysis results for alternative outcome definitions in the prescribing trends study.

	Probable AD	Any AD	Any dementia
Diagnoses	Probable AD	Probable AD Possible AD	Probable AD Possible AD Non-AD and mixed dementias
AChE inhibitors	Eligible: 10456 Treated: 5019 Joinpoint 1: Oct 2012 (Sep 2011 – Apr 2013) Joinpoint 2: May 2013 (Nov 2012 – Apr 2014)	Eligible: 21342 Treated: 6449 Joinpoint 1: Jun 1999 (Apr 1998 – Dec 2000) Joinpoint 2: Jun 2001 (Sep 2000 – Mar 2002)	Eligible: 38650 Treated: 9896 Joinpoint 1: Aug 2000 (Jun 1998 – Nov 2000) Joinpoint 2: Jan 2001 (Sep 2000 – Nov 2001)
NMDA receptor antagonists	Eligible: 9964 Treated: 1052 Joinpoint 1: Mar 2011 (Aug 2010 – Apr 2011) Joinpoint 2: Jun 2011 (Apr 2011 – Nov 2011)	Eligible: 18930 Treated: 1309 Joinpoint 1: Sep 2010 (Dec 2009 – Apr 2011) Joinpoint 2: Nov 2011 (Apr 2011 – Mar 2012)	Eligible: 35625 Treated: 1961 Joinpoint 1: Aug 2010 (Nov 2009 – Dec 2010) Joinpoint 2: Nov 2011 (Aug 2011 – Mar 2012)

Figure 6.4: Google trends data for news searches in England for the disease term 'Alzheimer's Disease' between January 2008 and December 2015.



This graph shows the interest in the disease term 'Alzheimer's disease' each month from January 2008 to December 2015 inclusive. Interest is given as a percentage scaled against peak popularity, which is represented as a value of 100% and occurred for the downloaded data in September 2012. Values of zero indicate insufficient data for that month.

6.5. Discussion

The first trend change for the proportion of patients with probable Alzheimer's disease receiving their first prescription for an acetylcholinesterase inhibitor occurred in October 2012 (CI: September 2011 to April 2013). There were two potentially relevant events that occurred at this time. Firstly, the patents expired on the three drugs in this class in 2012 – galantamine in January 2012; donepezil in February 2012; and rivastigmine in July 2012. Secondly, the Prime Minister's Dementia Challenge was launched in May 2012. It is likely that the reduction in cost of these drugs, which resulted from their patents expiring, in combination with increased awareness of dementia due to the Prime Minister's Dementia Challenge led to the observed increase in prescription rates. In addition to these factors, a large amount of literature concerning acetylcholinesterase inhibitors had been published ahead of the revisions to the NICE guidance in 2011. Although this is unlikely to have directly caused a change in prescribing, it could contribute to the long term steady increasing trend for prescribing these drugs. A systematic review, which covers the literature through November 2014 (i.e. after all join points identified in the analysis but 13 months before the end of the study), summarizes the literature available at that time. It shows several studies were published between 2003 and 2008 that suggested patients with mild to moderate Alzheimer's disease could benefit from acetylcholinesterase inhibitors with estimated "improvements on the order of 1.5 MMSE (30-point scale)". (185) Therefore, a potential effect of the literature on prescribing cannot be ruled out, even though the authors of the review questioned whether such an improvement was clinically meaningful when all the evidence was presented together. Further to the support from the literature, the Google trends data for news searches in England also suggested increased awareness around the time of this trend change. The interest for the search term 'Alzheimer's disease' was at its maximum in September 2012 (based on the data available from January 2008 to December 2015 inclusive), which could indicate interest among the public.

The second trend change in the Acetylcholinesterase inhibitor analysis occurred in May 2013 (CI: November 2012 to April 2014), less than a year after the initial change for this drug class and with overlapping 95% CIs. This change signals the start of the first decreasing trend in prescriptions. This is not unexpected as patent expiry may have led to a form of 'catch up prescribing', whereby people who were previously denied access to

the drug were granted access at this time because of its newly reduced cost. This would result in the apparent decreasing trend once 'catch up prescribing' was complete, which is suggested by the trend analysis but is not as clear when considering the raw data points. These results differ from the sensitivity analyses that considered relaxed diagnosis definitions, though the 'any Alzheimer's disease' and 'any dementia' analyses were in line with each other. This suggests that prescribing for patients with probable Alzheimer's disease was more consistent, as one might expect, across the study than for other groups. This could indicate that patients with dementias other than probable Alzheimer's disease, i.e. with unlicensed indications, were receiving these drugs and that their prescribing was subject to change over the period studied. Further to this, large increases in prescriptions are observed as the diagnosis definition is relaxed. This could provide further evidence for the possible unlicensed use of this drug class. The literature at that time also reflects ongoing discussion concerning the benefit of these drugs for indications other than Alzheimer's disease. For example, a 2012 review by Rodda and Carter discusses their use in vascular dementia, dementia with Lewy bodies and Parkinson's disease dementia. (186) Alternatively, it could be attributed to the fluctuating course of symptoms that some people with dementia experience or increased recognition of mixed diagnoses, where there is evidence of Alzheimer's disease in addition to other forms of dementia; both of which might lead to treatment changes.

The trend changes in the NMDA receptor antagonist analysis occurred in March 2011 (CI: August 2010 to April 2011) and June 2011 (CI: April 2011 to November 2011). In March 2011, NICE introduced guidelines that recommended the prescription of memantine for patients with moderate to severe Alzheimer's disease, or those people who could not tolerate acetylcholinesterase inhibitors. This replaced existing guidelines that restricted access to memantine to patients participating in clinical trials. It is therefore possible that these trend changes relate to the transition between the existing guidelines and those introduced in March 2011. In addition, the second highest peak in interest (88% of maximum interest) for the disease term 'Alzheimer's disease' in the Google trends data for news searches in England in January 2011. In this month, the 'Final Appraisal Determination on Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease' was released. NICE defined this document as 'the appraisal committee's final draft guidance about using a treatment or group of treatments in the NHS', which becomes guidance if not appealed. (187) The increase in

news searches around this time, and its alignment with the release of the final draft guidance, supports the idea of a transition in prescribing practice due to the NICE guidance. Finally, the evidence concerning the use of memantine is summarized in a technology appraisal conducted by NICE in 2011 to support their guidance. (168) This new information, based on several studies published prior to the trend change, might have played a role in changes to prescribing. Interestingly, neither of the trend changes in the NMDA receptor antagonist analysis align with those observed for the acetylcholinesterase inhibitors. This suggests that the NICE guidelines, which were implemented at the same time for both drug classes, may not have been as effective for acetylcholinesterase inhibitors as they appear to have been for NMDA receptor antagonists. This is likely because acetylcholinesterase inhibitors were available outside of clinical trials prior to the restrictive guidelines recommended in 2006. The sensitivity analyses conducted for the NMDA receptor antagonists were consistent with these results, regardless of the diagnosis definition used. The first of the joinpoints for all analyses occurred in the seven-month period between August 2010 and March 2011 and the second occurred in the six-month period between June 2011 and November 2011. This high level of consistency across diagnosis definitions indicates a clear pattern in prescribing, suggestive of a distinct change in practice. This provides additional support for the inferences concerning the impact of the 2011 NICE guidance on the NMDA receptor antagonist drug class.

Overall, analysis of both drug classes indicates that inclusion of dementia in QOF had no effect on prescribing trends and the other factors had mixed effects. NICE guidance on the prescribing of drugs for dementia aligned with trend changes for NMDA receptor antagonists but not acetylcholinesterase inhibitors. The guidance that had the noticeable effect was released in March 2011 and allowed the NMDA receptor antagonist, memantine, to be used outside of clinical trials. All other guidance for both this drug and acetylcholinesterase inhibitors, including that which recommended restricting access, did not align with trend changes. Government dementia strategies also appear to have had mixed results, with the Prime Minister's Dementia Challenge (launched May 2012) the only strategy to align with a trend change. Although this strategy is likely to have increased awareness of dementia around the time of the October 2012 trend change for acetylcholinesterase inhibitors, the patent expiry of the drugs in this class is more likely to be the cause. This is because it will have reduced the cost of these drugs and potentially

led to a surge in prescribing, such as that observed in the trend analysis. The events considered here highlight the many factors that may have influenced prescribing rates and the challenges in assessing the impact of a given event. Over the period studied, the proportion of patients receiving prescriptions increased, irrespective of changing guidelines and other initiatives. Furthermore, given the increase in diagnoses of dementia and, more specifically, Alzheimer's disease reported in the CPRD (Figure 6.1), the absolute number of prescriptions has increased considerably.

6.6. Strengths and limitations

The key strength of this study is the large sample of primary care data with prescribing information, provided by the CPRD. The data used in the present analysis contained 40,202 patients diagnosed with dementia in England up to 1st January 2016 (note that data is restricted to practices with a last data collection date in 2016). This included 10,651 patients with probable Alzheimer's disease and a further 12,167 patients with possible Alzheimer's disease. A further strength of the study is the long follow up of patients that allowed consideration of patients who did not receive immediate treatment. This is important as pharmacological interventions for Alzheimer's disease have historically considered severity as part of the prescribing decision and so there was likely to be a treatment delay after initial diagnosis for those presenting with mild disease.

The main limitation of this study is the likelihood of missed diagnoses. This is demonstrated within the dataset, as there were 1,231 patients receiving one of the treatments of interest who did not have any form of recorded dementia diagnosis. Missed diagnoses are likely to be due to: (1) outdated or non-specific diagnoses (i.e. type of dementia is not updated once established); (2) diagnoses received outside of primary care (i.e. from a specialist service); and (3) unrecorded diagnoses in primary care (i.e. a diagnosis is given but not added to a record). Missed diagnoses have been explored in sensitivity analyses by testing the sensitivity and specificity of the diagnosis definitions and by relaxing the diagnosis definition from 'probable Alzheimer's disease' to include other less certain codes for the disease and other types of dementia. Neither of these sensitivity analyses provided any cause for concern. A final limitation of this study is the difficulty in determining the lag time between an event and a trend change to assess the impact of the event. To allow for this, I focused on events considered to be of greatest

impact – i.e. changes at a national level – and so any effect associated with them is expected to be evident if present. However, this prevents me from covering all the factors that may have influenced prescriptions for drugs for dementia during the study period – for example, I cannot comment on all relevant articles published during this time.

6.7. Summary

This chapter demonstrates that prescription rates in England do not always respond to factors such as regulatory guidance, recommendations or patent expiry and, when they do, not necessarily in a predictable way. This has potential implications on how novel drugs in this therapy area, including repurposed drug candidates, are used. In addition, the analysis has provided insight into the factors that may have influenced prescription rates of drugs for dementia in England since their launch in 1997. This is essential for accurate assessment of the effectiveness of these treatments and to adjust for them in other forms of analyses, particularly those concerning treatments to alter the progression of dementia.

Chapter 7. Results: Assessment of the effects of antihypertensive drugs on dementia prevention using electronic health record data

7.1. Introduction

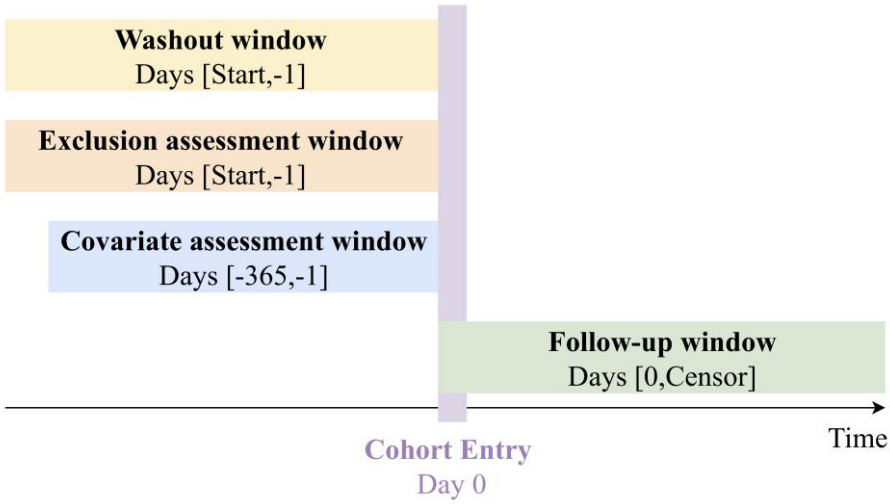
In Chapter 6, I demonstrated that the proportion of patients receiving drugs for dementia increased between June 1997 and December 2015, irrespective of changing guidelines and other initiatives. This highlights the huge unmet clinical need in this field. As discussed in Chapter 2, drug repurposing is one way in which this need could be met and antihypertensives have previously been highlighted as priority drug repurposing candidates. Therefore, in this chapter, I used instrumental variable analysis with physicians' prescribing preference as an instrument for exposure, which was introduced in Chapter 3, to assess whether antihypertensive drugs can be repurposed for dementia prevention. As in Chapter 6, I used data from the CPRD (Chapter 5). Several observational studies have investigated repurposing antihypertensives for dementia prevention previously. (12–19) However, as discussed in Section 1.2, these studies have used case-control designs with logistic regression and cohort designs with survival analysis, which may be subject to unmeasured or residual confounding and reverse causation. Therefore, the rationale for the study described in this chapter was to use instrumental variable analysis, which can be robust to confounding and reverse causation if certain assumptions are met, to provide new evidence about the potential effects of antihypertensives on risk of dementia.

7.2. Methods

7.2.1. Study design

I conducted a prospective new user cohort study in the CPRD, the design of which is detailed in Figure 7.1. (188) The CPRD is described fully in Chapter 5. The *a priori* protocol for this study was published prior to the study being conducted. (36)

Figure 7.1: Study design diagram for the assessment of the effects of antihypertensive drugs on dementia prevention using electronic health record data.



Study design diagrams, such as this, are a tool used in pharmacoepidemiology to describe the criteria applied to select patients for a study and the times when those criteria applied. The x-axis represents time and the boxes correspond to different restrictions. Patients entered the cohort when initially exposed to an antihypertensive drug (day 0, marked in purple). At this time, patients could not have been recorded as previously having accessed a drug of interest (washout window, marked in yellow) and could not have had a diagnosis preventing them from participating in the study (exclusion assessment, marked in orange), such as existing dementia, from 'Start' (i.e. the first day of data collection) to the day before they entered the cohort. Patients were also required to have 12 months of data prior to cohort entry to define baseline covariates (covariate assessment window, marked in blue). Patients were followed until they were censored (follow-up window, marked in green). This diagram explicitly shows that there is no overlap of follow-up with the exclusion and covariate assessments and washout period in this study.

Three amendments have been made to the study design since the publication of the protocol:

1. 'Centrally acting antihypertensives' and 'Loop diuretics' were not considered as exposures because the former is primarily used for acute events, while the latter is primarily used for heart failure.

2. The drug classes 'Potassium-sparing diuretics and aldosterone antagonists' and 'Thiazides and related diuretics' were combined into a single category titled 'Diuretics' as prescriptions for the former in the data extract were rare.

3. I used each drug class as the reference drug class in turn and presented all results in a matrix, instead of using beta-adrenoceptor blocking drugs as the reference drug class for all analyses.

7.2.2. Participants

Patients were included in the analysis if they were aged 40 years or over and received a first prescription for an antihypertensive drug class of interest. Follow-up was stopped at the earliest of: a dementia outcome; death; end of registration at a CPRD general practice; or the end of follow-up for this study (29th February 2016). Patients were excluded if they were of unknown gender; had less than 12 months of 'research quality' data prior to their first prescription (to improve the identification of baseline covariates); or were initially prescribed multiple antihypertensive drug classes of interest. I also excluded patients prescribed an antihypertensive before 1st January 1996, as 1996 was the first complete year that all drugs being considered were available.

7.2.3. Exposures

Seven antihypertensive drug classes based on the groupings in the British National Formulary were considered. (189) These were: alpha-adrenoceptor blockers, angiotensin-converting enzyme inhibitors, angiotensin-II receptor blockers, beta-adrenoceptor blockers, calcium channel blockers, diuretics (either 'thiazides and related diuretics' or 'potassium-sparing diuretics and aldosterone antagonists'), and vasodilator

antihypertensives. To mimic a randomised controlled trial, exposure to the drug classes was analysed in an intention-to-treat framework, i.e. based on the first prescription irrespective of subsequent switches to, or additions of, other antihypertensive drug classes. (73,190) The index date for each patient was the date they received their first prescription for an antihypertensive drug. Treatment switching was not modelled, as it was likely to be non-random and confounded by patients' unobservable characteristics.

7.2.4. Outcomes

There were four outcomes defined for this analysis: probable Alzheimer's disease, possible Alzheimer's disease, vascular dementia and other dementias. See Figure 5.1 for the decision tree used to determine dementia subtype. I also considered any dementia, which combined the dementia subtypes in a single outcome.

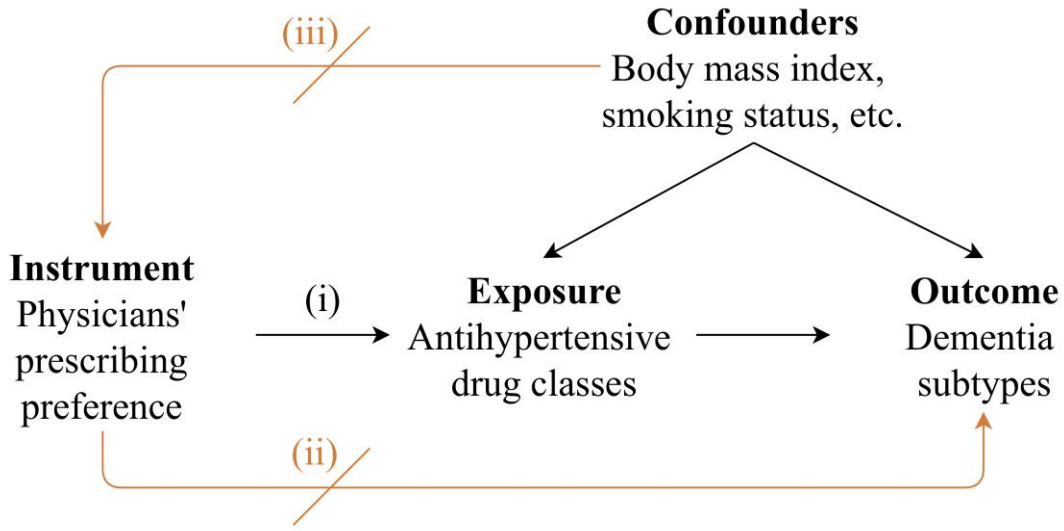
7.2.5. Covariates

The instrumental variable analysis was adjusted for prescription year only. This was necessary as the number of antihypertensive prescriptions in the CPRD varied by year and so may have influenced both the instrument-exposure and instrument-outcome associations. All other potential covariates were thought to influence the exposure-outcome association, but not the instrument-exposure or instrument-outcome associations, and so will be balanced across levels of the instrument if the instrument assumptions are met. The instrumental variable analysis was compared with a multivariable logistic regression analysis to assess the extent of confounding. The multivariable logistic regression analysis was adjusted for prescription year; sex; age at index; previous history of coronary heart disease, coronary-bypass surgery, or cerebrovascular disease; chronic disease; socioeconomic position; consultation rate; alcohol status; smoking status; and body mass index. All covariates were determined prior to index and are defined fully in Table 5.2.

7.2.6. Statistical methods

This study used instrumental variable analysis with physicians preferred antihypertensive drug class as an instrument to proxy for exposure, i.e. the actual drug class prescribed (Figure 7.2). Each drug class was used as the reference drug class for each of the other drug classes in a series of pairwise comparisons. Prescribing preference was derived from the prescriptions issued by the physician to their seven most recent patients who received an antihypertensive. (27,191) This resulted in an ordered categorical instrument indicating how many previous prescriptions the physician had issued for the drug class of interest over the reference drug class in the present pairwise comparison. I selected seven previous prescriptions to define this instrument as this is frequently used in the literature and is thought to balance instrument strength, which increases with additional prescriptions used, and recent prescribing trends, which are lost with additional prescriptions used. (25) The analysis used the ivreg2 package in Stata with 'robust' specified (to address arbitrary heteroskedasticity, i.e. non-normality in the outcome) and clustering by physician (to address both arbitrary heteroskedasticity and intra-group correlations between physicians). (192) To obtain a point estimate, I made a further assumption – in addition to the three standard instrument assumptions - of monotonicity (Section 3.2.2.2). That is, I assumed that physicians' preferred drug class had one of three effects: it increased the likelihood of exposure in everyone; it decreased the likelihood of exposure in everyone; or it had no effect in everyone. Consequently, the results were interpreted as the effect among patients whose prescription was affected by their physicians' preference (known as the local average treatment effect). (193) For each analysis, I present the partial F statistic to quantify and test the strength of the instrument-exposure association. I also present the results of endogeneity tests conducted using the option 'endog' in ivreg2. The analysis is presented in line with reporting guidelines. (194) All analyses were conducted in Stata (version 15MP) and R (version 3.4.4). (80,195) The code is available from GitHub: https://github.com/venexia/repurposing-antihypertensives-dementia.

Figure 7.2: Illustration of the instrumental variable analysis model used for the assessment of the effects of antihypertensive drugs on dementia prevention using electronic health record data.



As detailed in Section 3.2, instrumental variable analysis requires that the instrument: (i) be associated with the exposure of interest; (ii) affect the outcome only through its effect on the exposure of interest; and (iii) have no common causes with the outcome. To obtain a point estimate for this analysis, I also make a fourth assumption of monotonicity (Section 3.2.2.2). The measured confounders of this analysis are listed in the section 'covariates', however there is also likely to be unmeasured confounders of the exposure-outcome association hence warranting the use of this method.

7.2.7. Assessment of bias

To assess bias, I constructed bias scatter plots for each outcome. These plots compare the association of each covariate with the exposure (obtained from multivariable linear regression analysis) and the instrument (obtained from instrumental variable analysis). (196,197) If all points on these plots were on the x-axis then this would indicate that the instrumental variable analysis would be less biased than multivariable linear regression analysis. Points above the x-axis but below the x=y line would indicate bias in both analyses that is greater in the multivariable linear regression analysis estimate, while points above the x=y line would indicate bias in both analyses that is greater in the instrumental variable analysis estimate. Points on the x=y line would indicate that the bias is equal for the two analyses. Any covariates found to be as, or more, biased for the instrumental variable analysis (i.e. on or above the x=y line) were adjusted for in a sensitivity analysis.

7.2.8. Sensitivity analyses

Beta-adrenoceptor blockers can be prescribed in low doses for the treatment of anxiety. (198). However, to be a suitable comparator, they must be prescribed for the treatment of hypertension. I therefore tested the effect of removing patients thought to be receiving these drugs for anxiety in two ways. Firstly, I did the analysis without people who both received a drug class of interest and had a Read code indicating anxiety, or other neurotic, stress-related and somatoform disorders in the same consultation (using a previously published code list). (199) Secondly, I did the analysis without people whose dose was in the bottom 25% for their index drug class.

Differential prescribing occurs in women of child bearing age due to risks associated with some antihypertensives during pregnancy. (58) As participants can enter this study at the age of 40, this might affect the youngest members of the cohort. I therefore conducted a sensitivity analysis restricted to patients aged 55 and over at index. This age threshold is currently being used in the RADAR trial for similar reasons. (200)

7.3. Results

7.3.1. Patient characteristics

A total of 849,378 patients, with a total follow up of 5,497,266 patient years, met the criteria for the analysis. Figure 7.3 outlines patient attrition and Table 7.1 presents patient characteristics of those remaining in the study. (201–203) The full cohort had a median age of 61 (interquartile range: 51 to 71) at index date and a median follow-up of 5.8 years (interquartile range: 2.6 to 9.8). Of the 849,378 patients, 410,805 (48%) had complete covariate information. This subset of patients was used when comparing instrumental variable and multivariable logistic regression analyses. The subset had a median age of 61 (interquartile range: 51 to 71) at index date, and median follow-up of 5.6 years (interquartile range: 2.5 to 9.5). Incomplete covariate information was mainly due to missing values for the Index of Multiple Deprivation, which was used to adjust for socioeconomic position, as this measure is only available for patients in English practices. One notable feature of the patient characteristics is that 97% of patients receiving alpha-adrenoceptor blockers and 99.3% of patients receiving vasodilator antihypertensives were men - this difference persists regardless of the age at first prescription (Table 7.2).

7.3.2. Alzheimer's disease

Figure 7.4 shows the results for probable and possible Alzheimer's disease respectively. The results suggested that beta-adrenoceptor blockers were protective for both probable and possible Alzheimer's disease when compared with other drugs. For example, beta-adrenoceptor blockers were estimated to result in 8 (95% CI: 3 to 12; p-value=3.1×10^{-3}) fewer cases of probable Alzheimer's disease and 9 (95% CI: 4 to 13; p-value $=1.3 \times 10^{-4}$) fewer cases of possible Alzheimer's disease per 1000 people treated when compared with alpha-adrenoceptor blockers.

Figure 7.3: Patient attrition for the assessment of the effects of antihypertensive drugs on dementia prevention using electronic health record data.

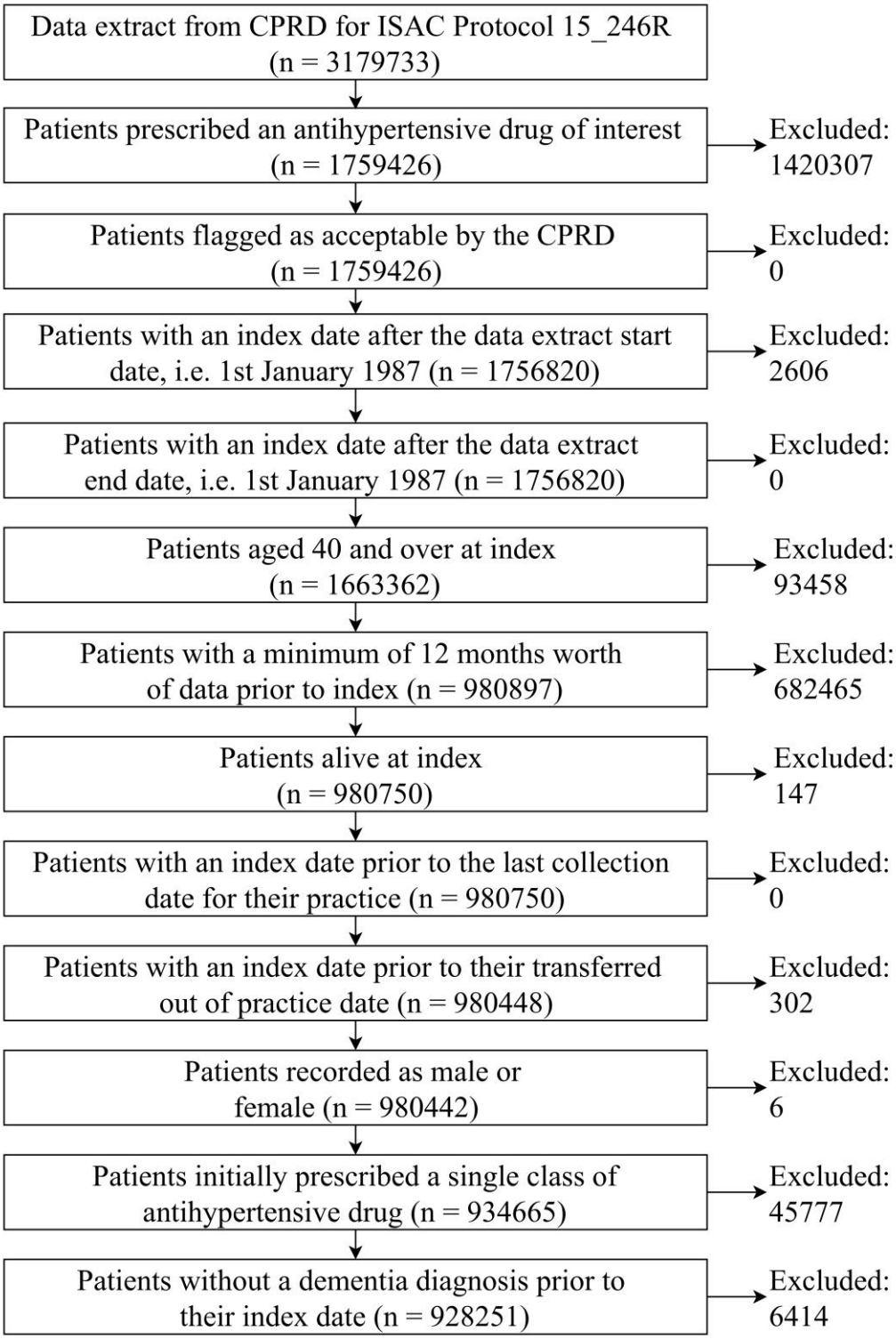


Table 7.1: Patient characteristics for the assessment of the effects of antihypertensive drugs on dementia prevention using electronic health record data.

	Alpha-adrenoceptor blockers	Angiotensin-II receptor blockers	Angiotensin-converting enzyme inhibitors	Beta-adrenoceptor blockers	Calcium channel blockers	Diuretics	Vasodilator antihypertensives	Whole sample
N	67360	14717	195891	240864	139730	180946	9870	849378
Median year of first prescription	2008	2005	2007	2005	2008	2003	2008	2006
Male sex	97.0% (65365)	55.3% (8141)	58.0% (113667)	43.2% (104096)	49.2% (68739)	36.0% (65177)	99.3% (9796)	51.2% (434981)
Median age at first prescription	65	59	59	55	64	66	57	61
Previous history of coronary artery disease	0.2% (129)	0.6% (85)	0.8% (1536)	0.9% (2056)	0.4% (562)	0.1% (203)	0.1% (11)	0.5% (4582)
Previous history of coronary-bypass surgery	0.3% (193)	0.3% (45)	0.5% (946)	0.5% (1262)	0.3% (418)	0.1% (265)	0.1% (14)	0.4% (3143)
Previous history of cerebro-vascular disease	2.0% (1319)	2.1% (311)	3.0% (5813)	1.4% (3387)	2.3% (3194)	2.8% (5090)	0.7% (73)	2.3% (19187)
At least one comorbidity on the Charlson index[a]	36.8% (24817)	42.4% (6238)	50.8% (99492)	26.0% (62604)	38.7% (54081)	36.0% (65212)	42.6% (4207)	37.3% (316651)
Median IMD 2010 score[b]	8	8	9	9	9	9	8	9
Mean annual consultation rate (SD)	5.6 (5.4)	6.1 (6.3)	6.1 (6.0)	5.8 (5.3)	5.9 (5.8)	6.0 (5.6)	5.5 (5.1)	5.9 (5.7)
Ever drinker[c]	89.2% (60070)	85.2% (12538)	85.6% (167636)	86.1% (207457)	84.5% (118104)	84.3% (152473)	91.8% (9059)	85.6% (727337)
Ever smoker[d]	54.5% (36691)	52.5% (7729)	53.8% (105401)	54.3% (130894)	53.3% (74540)	55.2% (99793)	57.6% (5688)	54.2% (460736)
Mean body mass index (SD)[e]	26.5 (4.2)	28.6 (5.7)	29.0 (5.9)	26.6 (5.0)	27.5 (5.4)	27.5 (5.5)	27.3 (4.4)	27.5 (5.4)

SD: standard deviation

(a) The Charlson index is a classification of 17 chronic diseases, including cancer and arthritis, which may alter mortality risk.

(b) Index of Multiple Deprivation (IMD) 2010 score is a proxy for socioeconomic position that is measured as 'twentiles' with 1 indicating the least deprived and 20 indicating the most deprived. IMD 2010 score was missing for 38.6% (328,233 patients) of the whole sample.

(c) Alcohol status was missing for 15.6% (132,387 patients) of the whole sample. For the purposes of this table, it has been classified as 'ever' (i.e. former or current) vs 'never'.

(d) Smoking status was missing for 6.4% (54,447 patients) of the whole sample. For the purposes of this table, it has been classified as 'ever' (i.e. former or current) vs 'never'.

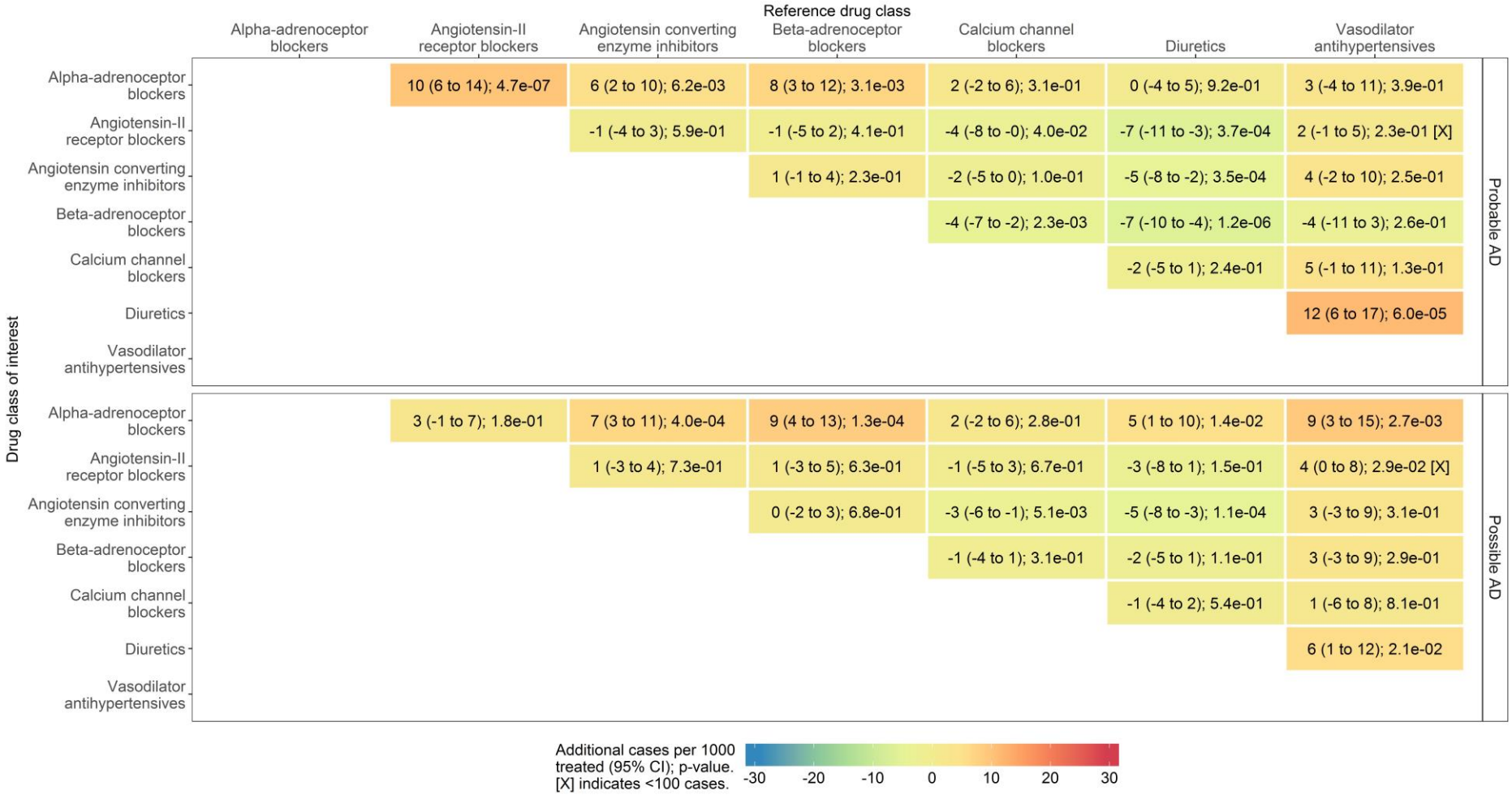
(e) Body mass index, or a calculated body mass index from height and weight measurements, was missing for 15.7% (128,830 patients) of the whole sample.

Table 7.2: Exposure by age and sex for the assessment of the effects of antihypertensive drugs on dementia prevention using electronic health record data.

Drug class / Age group	40-49	50-59	60-69	70-79	80-89	90-99	100+	Total
Alpha-adrenoceptor blockers	5,979 (93.4)	15,289 (97.1)	22,482 (98.0)	16,141 (97.5)	6,653 (96.3)	806 (94.7)	10 (100.0)	67,360 (97.0)
Angiotensin-II receptor blockers	42,697 (62.1)	57,932 (62.9)	45,726 (60.0)	32,124 (50.4)	14,881 (42.1)	2,482 (32.6)	49 (20.4)	195,891 (58.0)
Angiotensin converting enzyme inhibitors	2,994 (63.1)	4,366 (61.5)	3,742 (55.2)	2,479 (44.7)	1,010 (36.2)	125 (21.6)	1 (0.0)	14,717 (55.3)
Beta-adrenoceptor blockers	77,343 (39.3)	69,672 (45.7)	49,775 (48.2)	30,948 (43.0)	11,579 (35.7)	1,522 (26.2)	25 (20.0)	240,864 (43.2)
Calcium channel blockers	16,569 (50.6)	31,237 (55.8)	45,022 (53.8)	32,318 (42.7)	12,832 (34.9)	1,721 (23.7)	31 (16.1)	139,730 (49.2)
Diuretics	22,189 (33.3)	38,671 (39.2)	47,552 (40.9)	45,820 (34.2)	22,961 (28.9)	3,660 (23.1)	93 (11.8)	180,946 (36.0)
Vasodilator anti-hypertensives	2,440 (99.2)	3,381 (99.5)	3,033 (99.7)	907 (98.6)	103 (91.3)	5 (40.0)	1 (0.0)	9,870 (99.3)
Total	170,211 (48.5)	220,548 (55.2)	217,332 (56.3)	160,737 (47.7)	70,019 (40.5)	10,321 (31.5)	210 (19.5)	849,378 (51.2)

The number in brackets is the percentage of patients who are male.

Figure 7.4: Instrumental variable estimates for the risk of probable and possible Alzheimer's disease using electronic health record data.



F greater than 4708 for all analyses.

7.3.3. Non-Alzheimer's disease dementias

Figure 7.5 shows the results for vascular and other dementias respectively. The magnitude of the differences between drug classes is smaller for these outcomes. However, vasodilator antihypertensives were suggested to be protective with an estimated 5 (95% CI: 0 to 9; p-value=0.04) fewer cases of vascular dementia and 6 (95% CI: 1 to 11; p-value=0.02) fewer cases of other dementias per 1000 people treated when compared with calcium channel blockers. Angiotensin-II receptor blockers were also indicated to be protective for vascular dementia with an estimated 7 (95% CI: 4 to 10; p-value=1.4×10^{-5}) fewer cases of vascular dementia per 1000 people treated when compared with alpha-adrenoceptor blockers.

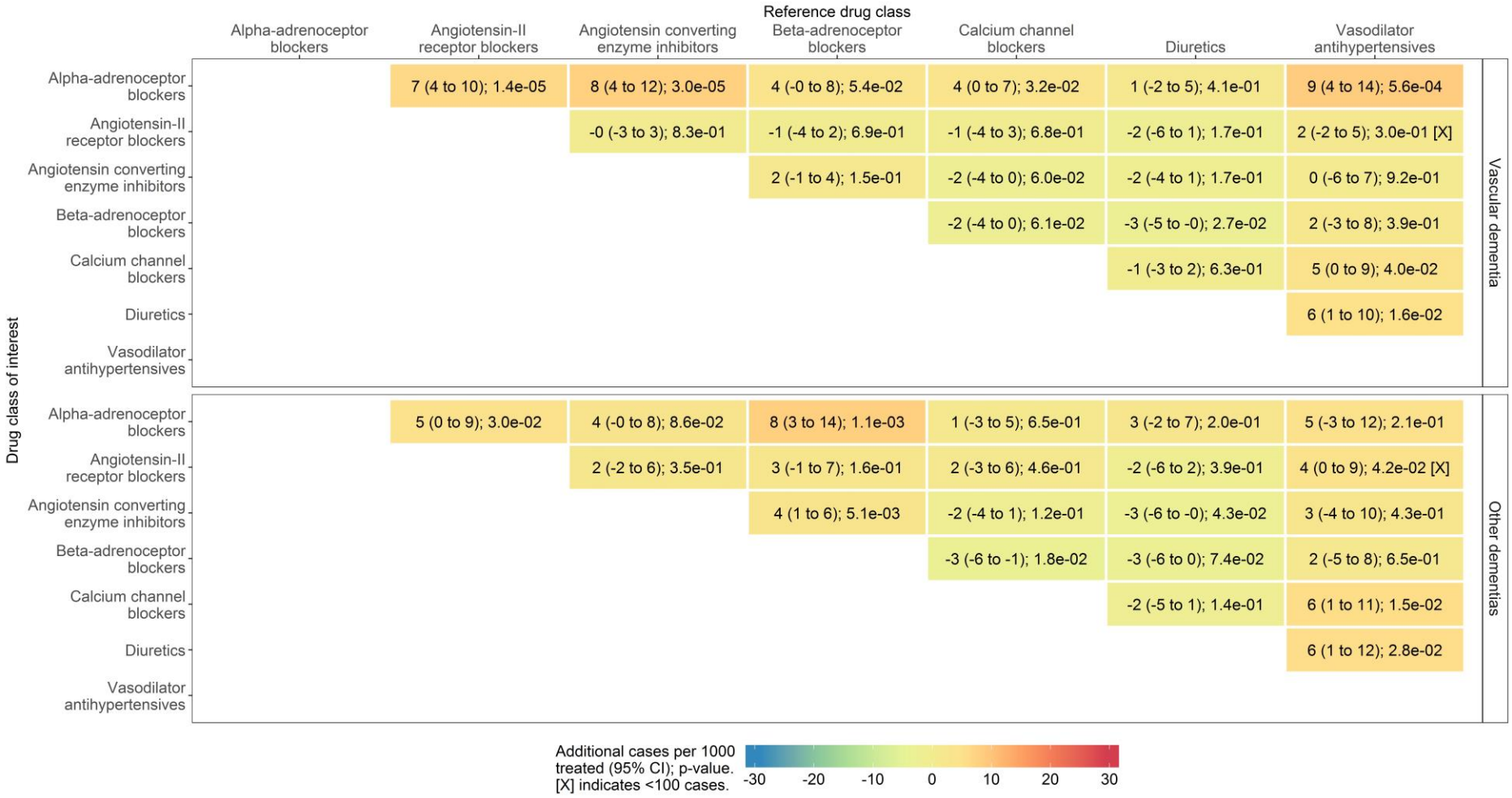
7.3.4. Any dementia

Figure 7.6 shows the results for any dementia. These results reflected the dementia subtype analyses and emphasised the effects observed, perhaps due to the increased sample size. For example, beta-adrenoceptor blockers were estimated to result in 28 (95% CI: 19 to 38; p-value=5.2×10^{-9}) fewer cases per 1000 people treated compared with alpha-adrenoceptor blockers. Meanwhile, vasodilator antihypertensives were estimated to result in 27 (95% CI: 17 to 38; p-value=4.4×10^{-7}) fewer cases per 1000 people treated compared with diuretics.

7.3.5. Comparison with multivariable logistic regression

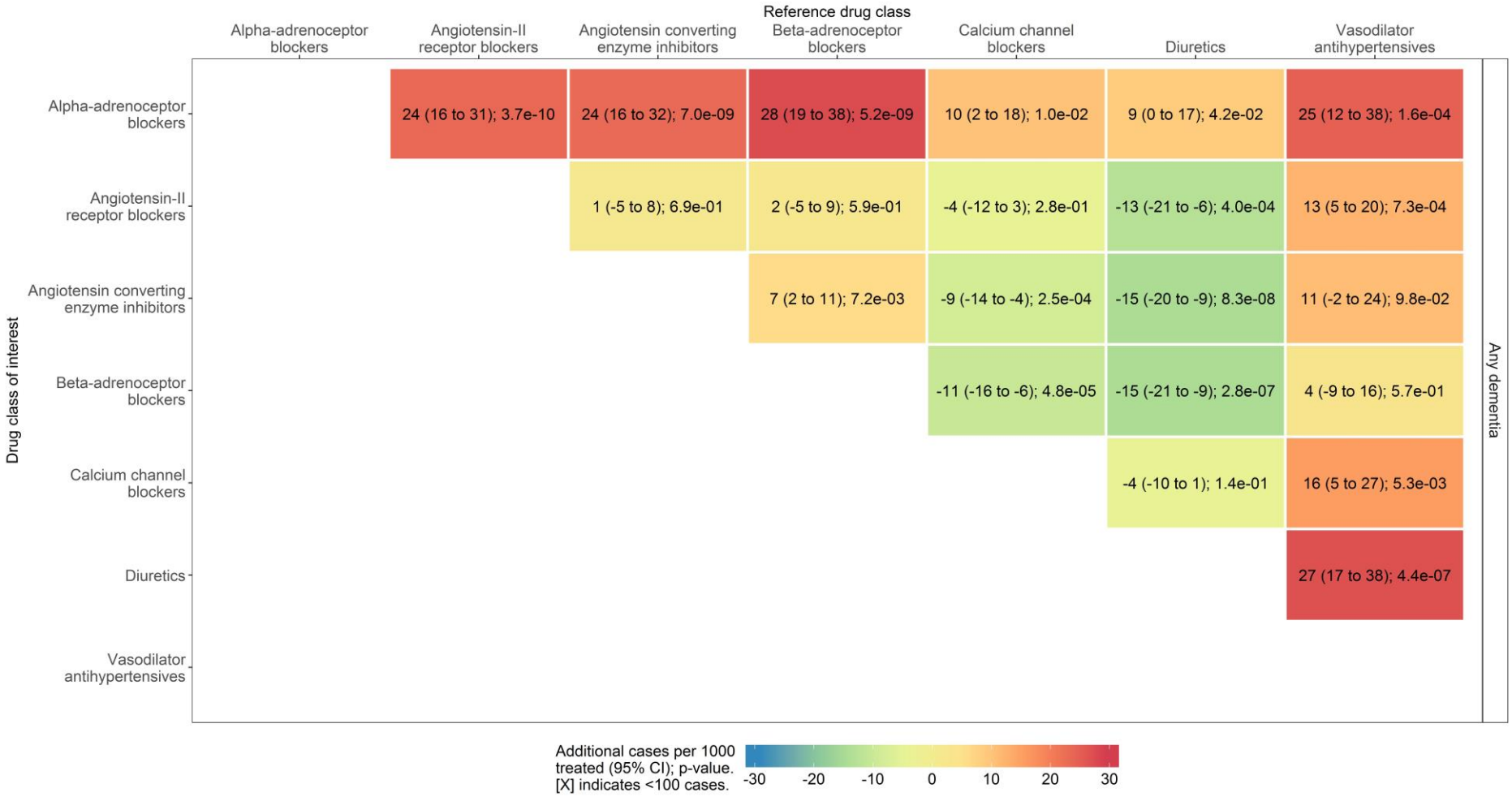
The results of the multivariable logistic regression are provided in Figure 7.7. Endogeneity tests indicated evidence to reject the null that the exposure was endogenous, indicating a difference between the instrumental variable analysis and ordinary least squares results, for a small number of the analyses run. Most of these analyses considered alpha-adrenoceptor blockers as the drug class of interest. The complete results, including the endogeneity test results and first stage regression results from the instrumental variable analysis, can be downloaded from the following link: http://rebrand.ly/repurposing-antihypertensives-dementia-supplement.

Figure 7.5: Instrumental variable estimates for the risk of non-Alzheimer's disease dementia using electronic health record data.



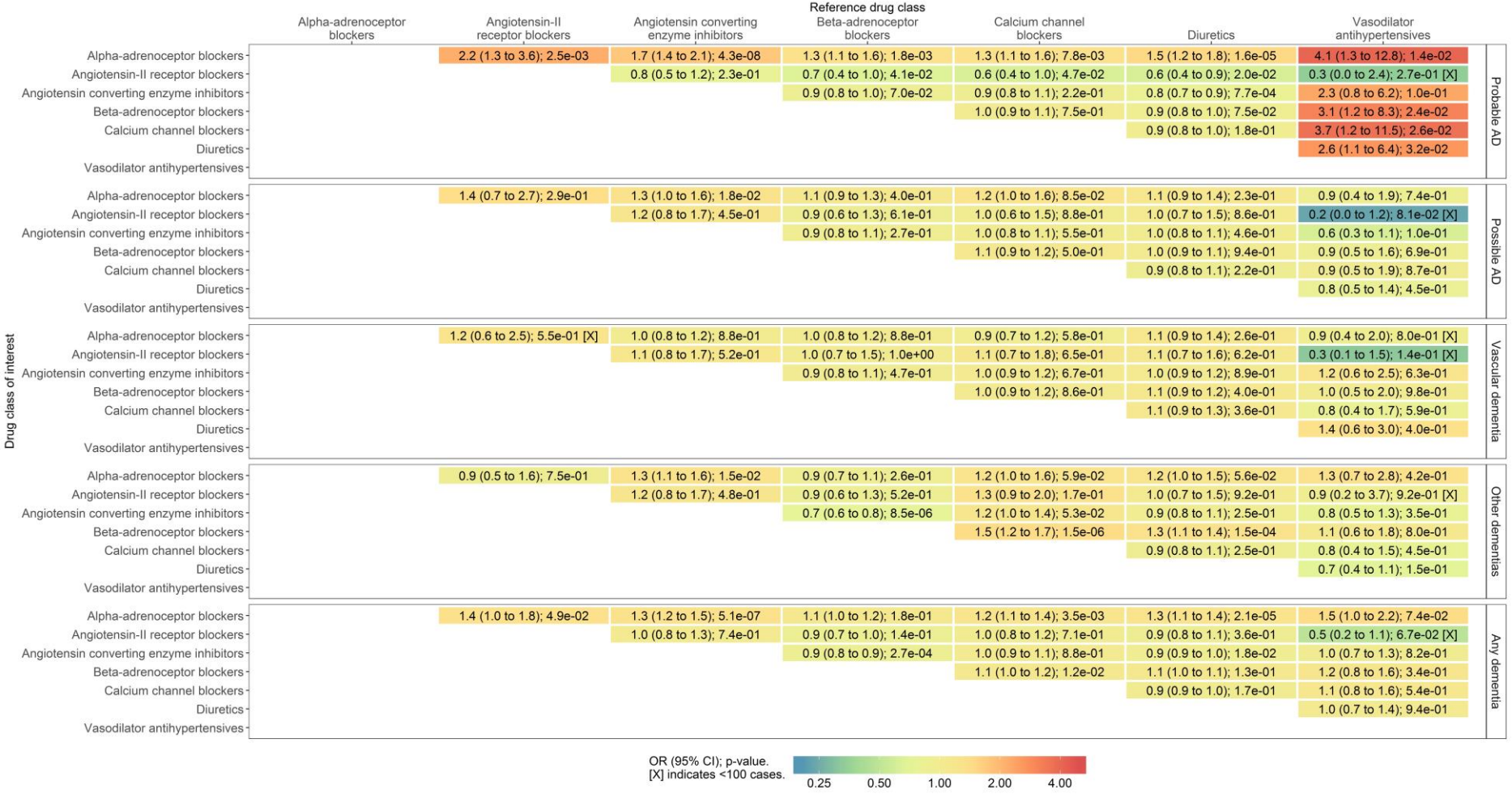
F greater than 4702 for all analyses.

Figure 7.6: Instrumental variable estimates for the risk of any dementia using electronic health record data.



F greater than 4876 for all analyses.

Figure 7.7: Multivariable logistic regression results for all dementia outcomes using electronic health record data.



7.3.6. Assessment of bias

Bias scatter plots were used to assess bias among the subset of patients with complete covariate information (Figure 7.8). The bias term was larger in the instrumental variable analysis, compared to the multivariable linear regression analysis, for socioeconomic position only. Bias terms were equally biased for body mass index, chronic disease, sex and age. These covariates, including socioeconomic position, were adjusted for in sensitivity analyses and were mostly found to produce consistent results with the main analysis (Appendix C). The exception was results concerning diuretics and beta-adrenoceptor blockers after adjustment for age. These drug classes have the oldest and youngest median ages at index respectively (Table 7.1), which may explain why they were most effected by the adjustment.

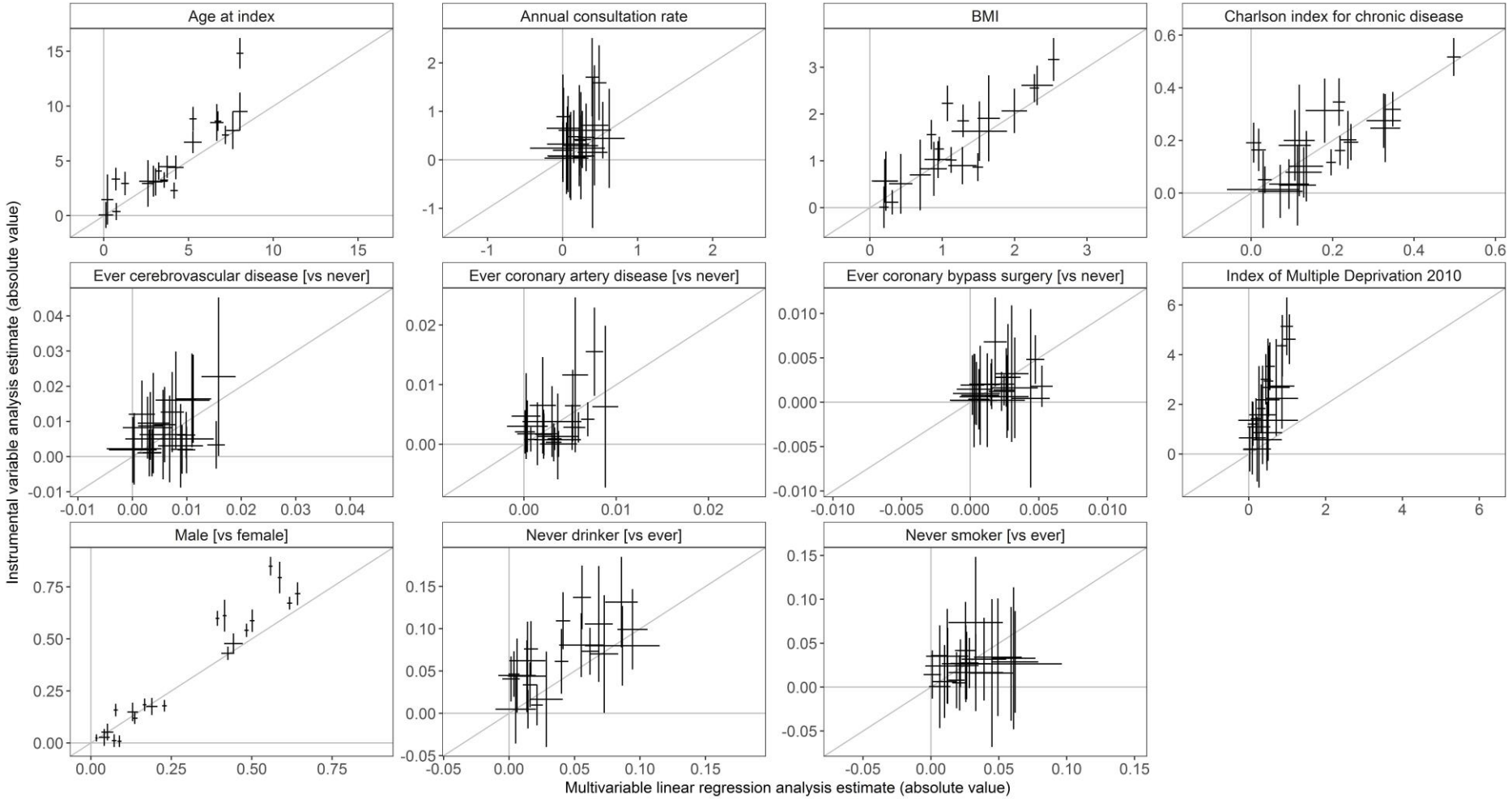
7.3.7. Sensitivity analyses

There was minimal effect of removing those diagnosed with anxiety in the same consultation from the analysis (Figure 7.9). Similarly, I observed little difference after removing those who received a low dose initial prescription though there was a lack of power for some analyses (Figure 7.10). Finally, restricting the analysis to patients aged 55 and over at index did not change the direction of effect for the results however, several effects failed to exclude the null after being subject to this restriction (Figure 7.11).

7.4. Discussion

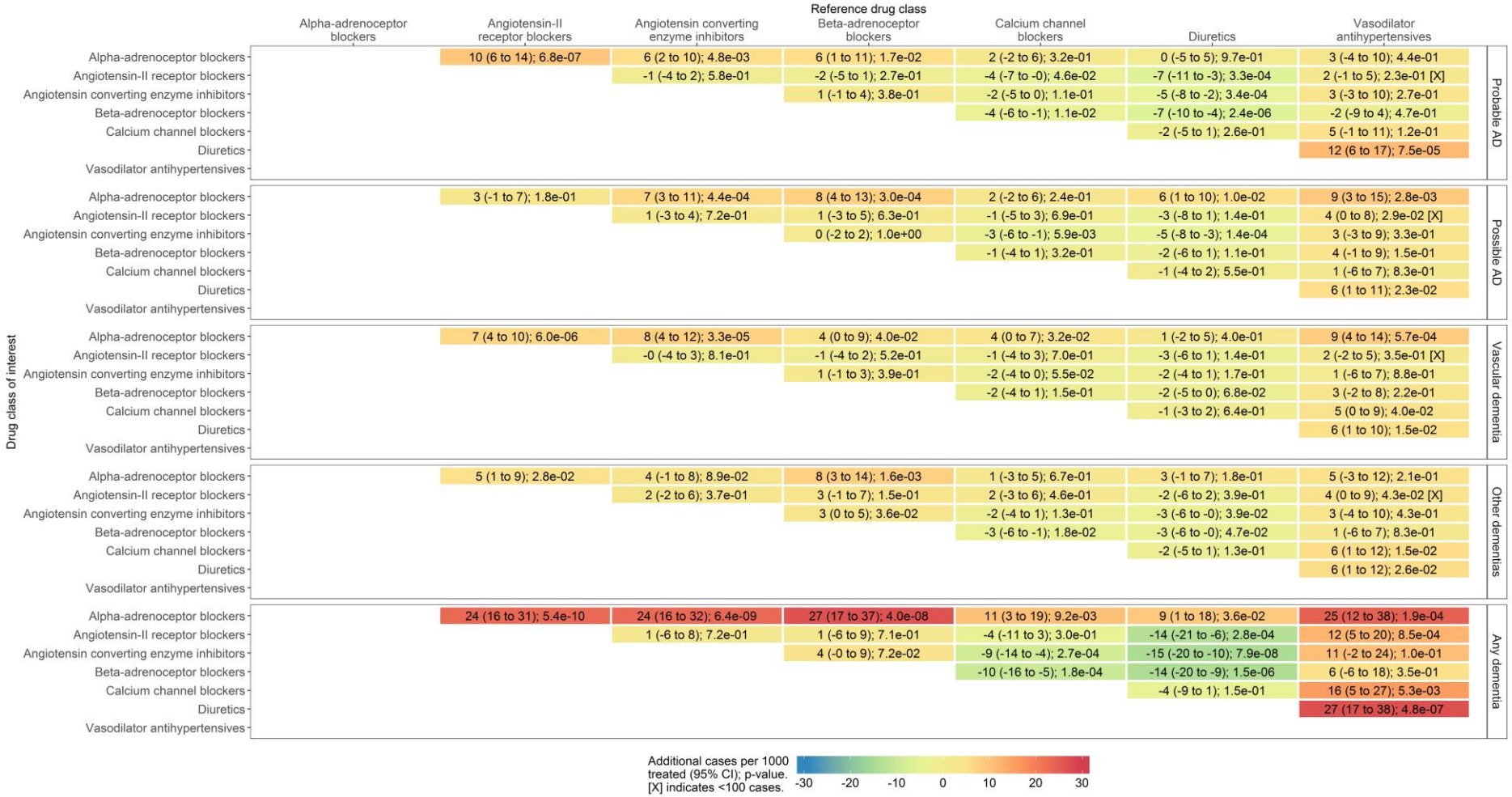
Beta-adrenoceptor blockers and vasodilator antihypertensives reduced risk of probable and possible Alzheimer's disease, vascular dementia, other dementias, and any dementia when compared with other antihypertensive drug classes. On the contrary, diuretics and alpha-adrenoceptor blockers increased risk of dementia outcomes when compared with other antihypertensive drug classes. The results concerning beta-adrenoceptor blockers and diuretics may be biased by age; however, this bias is no more extreme than that observed for multivariable linear regression. This study does not explore the effect of antihypertensives treatment compared to non-treatment on risk of dementia, which the meta-analysis of randomized controlled trials discussed in Section 2.7 suggests has a relative risk of 0.84 (95% CI: 0.69 to 1.02; p-value=0.10). (12)

Figure 7.8: Bias scatter plot for covariates in the any dementia analysis using electronic health record data.



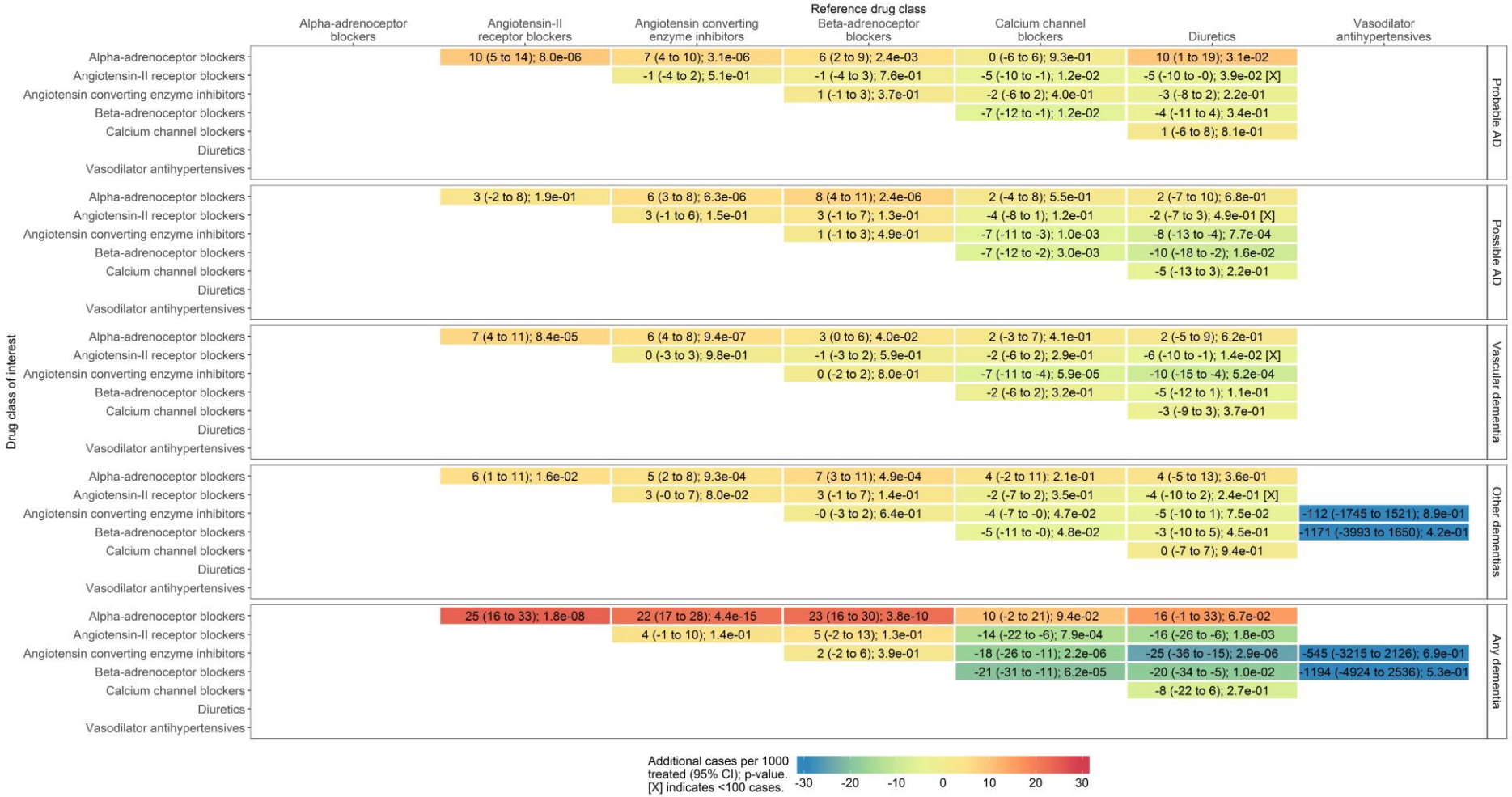
Each point on a scatter plot represents an individual analysis with the outcome 'any dementia'. This plot is representative of the results obtained for all outcomes.

Figure 7.9: Instrumental variable estimates for all dementia outcomes without patients diagnosed with anxiety in the same consultation using electronic health record data.



F greater than 4705 for all analyses.

Figure 7.10: Instrumental variable estimates for all dementia outcomes without patients who had a low dose initial prescription using electronic health record data.



F greater than 13 for all analyses. Missing cells due to insufficient sample size to run the analysis.

Figure 7.11: Instrumental variable estimates for all dementia outcomes for patients aged 55 and over at index using electronic health record data.

F greater than 1956 for all analyses.

Although there are currently some trials underway (Table 2.1), there have been no randomized controlled trials published to date that have directly compared antihypertensive drug classes to each other for the prevention or treatment of Alzheimer's disease. However, as discussed in Section 2.7, a recent meta-analysis by Larsson et al identified seven prospective observational studies on this topic. (12–19) Relative to other antihypertensive drug classes: one study (of three) suggested angiotensin-converting enzyme inhibitors were protective (13–15); three studies (of four) suggested angiotensin-II receptor blockers were protective (15–18); and one study (of one) suggested calcium channel blockers were protective (19). In contrast, this analysis suggested beta-adrenoceptor blocker and vasodilator antihypertensives were among the most protective drug classes when compared with other antihypertensives. Also mentioned in Section 2.7, is the study by Barthold et al comparing Alzheimer's disease incidence between users of renin-angiotensin system acting drug classes (angiotensin-converting enzyme inhibitors and angiotensin-II receptor blockers) and non-renin-angiotensin system acting drug classes (beta-adrenoceptor blockers, calcium channel blockers, loop diuretics, and thiazide-like diuretics) across sex, race, and ethnic groups in the USA. (62) Barthold et al found that angiotensin-II receptor blockers may reduce the risk of Alzheimer's disease in white and black women and white men. The study presented in this chapter is in a population of mainly white men and women and did not find such a clear distinction between angiotensin-II receptor blockers and non-renin-angiotensin system acting drugs. As already highlighted, the major difference between this observational study and those previously conducted is the statistical methods used. When the analysis assumptions are met, instrumental variable analysis should not be subject to unmeasured confounding, which may affect other types of analysis.

Two of the existing studies also made use of the CPRD. The first, by Davies et al, investigated the effects of angiotensin converting enzyme inhibitors and angiotensin-II receptor blockers, compared with other antihypertensives, on various dementia outcomes. (15) There is a small overlap between the present study and Davies et al, which I estimate to be 5.2% at most (48,363 new users of antihypertensives in Davies et al vs 849,378 new users of antihypertensives in the present study). The second, by Goh et al, compared the effects of angiotensin-II receptor blockers and angiotensin-converting enzyme inhibitors against each other in relation to dementia as a single outcome. (18) As they did not consider other antihypertensive drug classes as an exclusion criterion, they

had a much larger sample of 426,089 participants (as opposed to 221,421 participants in my study) exposed to angiotensin-II receptor blockers and angiotensin-converting enzyme inhibitors. This made it difficult to calculate the overlap as many of these patients are likely to have been exposed to other antihypertensives. However, there were 50,404 participants assigned to the drug classes angiotensin-II receptor blockers and angiotensin-converting enzyme inhibitors in this analysis that were not present in the Goh et al study as they received their initial treatment after 2010, i.e. after the final data extract for the Goh et al study. Despite the overlap of some of the data used in the present study with these studies in the literature, the study design and analysis differ considerably between them. One important distinction is that I systematically investigated all the major classes of antihypertensive drugs instead of focusing on those related to the renin-angiotensin system. This allowed me to consider multiple comparisons and potentially identify beneficial and harmful effects of non-renin-angiotensin system acting drug classes.

7.5. Strengths and limitations

The key strength of this study was the large cohort of patients (consisting of 849,378 patients with 5,497,266 patient years of follow-up) that would not be achievable in a randomized controlled trial. The size of this study meant there was ample power to detect even small differences between the drug classes of interest. There was also data on both male and female patients unlike some of the larger existing studies. (16) In addition, I used instrumental variable analysis, which should not be subject to unmeasured confounding when the assumptions hold, and an active comparator design, whereby antihypertensive drug classes were compared with other antihypertensive drug classes, to ensure patients were comparable. (73,188) These analysis features were important in the present study due to the risk of confounding. Despite finding some evidence of bias, the sensitivity analyses showed the effect on the results to be minimal with only minor changes in magnitude for most estimates. This includes the potential bias due to socioeconomic position, which was deemed the most extreme in my assessment. The only concern was potential confounding by age for results relating to beta-adrenoceptor blockers and diuretics, however this bias was no more extreme than that observed in the multivariable linear regression.

A limitation of the present study is that I cannot prove that the instrumental variable assumptions hold. The only assumption that can be empirically tested is the first, namely that the instrument is associated with the rates of prescribing (IV1: relevance –Section 3.2.1). The proposed instruments had a minimum F statistic of 4702 in the main analyses, demonstrating they strongly associated with the exposure. The study may also have misclassified the exposure due to the use of the intention-to-treat framework, which defines exposure based on the first treatment prescribed. However, the benefits of this approach – such as preserving sample size and replicating 'real world' prescribing – outweigh the concerns. A further limitation, as for the study in Chapter 6, is that this study may have misclassified outcomes, which can occur when a diagnosis is not updated or recorded accurately in primary care records. Again, I took steps to overcome this by considering 'probable' and 'possible' definitions for Alzheimer's disease, the most common form of dementia. I also included an 'any dementia' outcome that should not be affected by the difficulties of determining subtype.

7.6. Summary

This study provided new evidence about the potential effects of antihypertensives on risk of dementia through the novel application of instrumental variable analysis to this research question. I found small differences in drug class effects on risk of dementia outcomes. However, I showed the magnitude of the differences between drug classes is smaller than many observational studies, which may have been subject to unmeasured or residual confounding, have previously reported. These results will be discussed further in the context of the existing literature in Chapter 9, which will also cover the implications of this research.

Chapter 8. Results: Assessment of the effects of antihypertensive drugs on dementia prevention using genetic data

8.1. Introduction

As mentioned in Chapter 2 and demonstrated in Chapter 6, there is a huge unmet clinical need for novel drugs for dementia. In Chapter 7, I used instrumental variable analysis with physicians' prescribing preference as an instrument for exposure to investigate whether antihypertensives could be repurposed for dementia prevention. In this chapter, I used genetic instruments in a Mendelian randomization framework, as proposed in Chapter 4, to assess this same question. This genetically-informed method used different data to my previous assessment and was therefore subject to different biases. Consequently, it provides another strand of evidence that can ultimately be considered in a triangulation framework to obtain a reliable answer concerning the potential repurposing of antihypertensives for dementia. (20) Mendelian randomization has previously been used to study the relationship between genetically-predicted blood pressure and Alzheimer's disease but it has not been used to estimate the effects of the twelve most common antihypertensive drug classes on Alzheimer's disease. (204–206) In this chapter, I used SNPs as genetic instruments, selected to mimic the action of the protein targets of antihypertensive drug classes, in a two-sample Mendelian randomization analysis of systolic blood pressure on Alzheimer's disease. The rationale behind this chapter was to understand if there were differences between specific antihypertensive drug classes on Alzheimer's disease risk, which could inform the prioritization of repurposing candidates, and provide evidence at the drug class level that could be triangulated with that from other sources. (20)

8.2. Methods

8.2.1. Study design

I conducted a two-sample Mendelian randomization analysis using summary data on SNPs from GWAS, as described in Section 4.4. I identified SNPs to proxy exposure to an antihypertensive drug on the basis that they mimicked the action of that drug on their molecular targets. For example, angiotensin-converting enzyme inhibitors work by

inhibiting the enzyme angiotensin-converting enzyme. I therefore selected SNPs in the angiotensin-converting enzyme gene to use as a genetic proxy for this drug class. Effect sizes for these SNPs were then extracted from a GWAS of systolic blood pressure to estimate the instrument-exposure association. (207) The instrument-outcome association was estimated using the effect sizes for these same SNPs from a GWAS of Alzheimer's disease. (208) All data used were publicly available and mostly obtained from European ancestry populations.

8.2.2. Systolic blood pressure phenotype

The systolic blood pressure phenotype was defined using a GWAS of the UK Biobank cohort. (207) UK Biobank consists of 503,317 Caucasian people from the UK, aged between 38 years and 73 years. (209,210) The GWAS was based on 317,754 of the participants.

8.2.3. Alzheimer's disease phenotype

The Alzheimer's disease phenotype was defined using the International Genomics of Alzheimer's Project GWAS Stage 1 results. These data were from a meta-analysis of 17,008 Alzheimer's disease cases and 37,154 controls of European ancestry. (208).

8.2.4. Instrument selection

I identified twelve antihypertensive drug classes in the British National Formulary. (189) They were: adrenergic neurone blocking drugs; alpha-adrenoceptor blockers; angiotensin-converting enzyme inhibitors; angiotensin-II receptor blockers; beta-adrenoceptor blockers; calcium channel blockers; centrally acting antihypertensive drugs; loop diuretics; potassium-sparing diuretics and aldosterone antagonists; renin inhibitors; thiazides and related diuretics; and vasodilator antihypertensives. Using the drug substance information, I was able to identify pharmacologically active protein targets and the corresponding genes in the DrugBank database (version 5.1.1). (211) I then identified SNPs to instrument each target using the Genotype-Tissue Expression (GTEx) project data (Release V7; dbGaP Accession phs000424.v7.p2), which contains expression quantitative trait loci analysis of 48 tissues in 620 donors. (212) The full GTEx dataset,

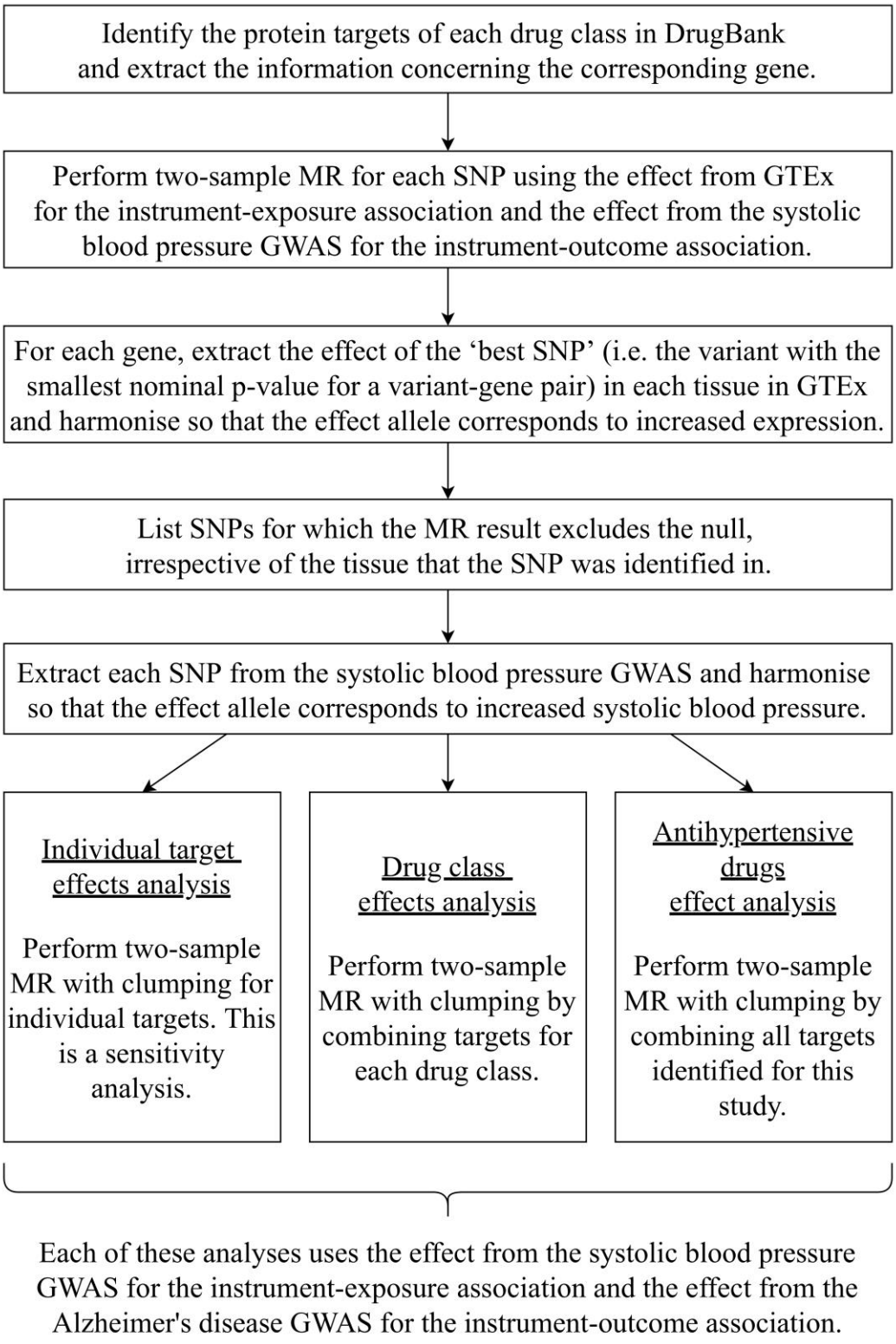
which consists of 714 donors, is 65.8% male and 85.2% white. SNPs marked as the 'best SNP' for the gene (defined by GTEx as the variant with the smallest nominal p-value for a variant-gene pair) in any tissue were selected for analysis. I considered SNPs identified in any tissue because: (i) as described below, these SNPs are validated by estimating their effect on systolic blood pressure prior to the main analysis so it was better to be inclusive at this stage; and (ii) it was difficult to determine the single most relevant tissue for systolic blood pressure.

To validate the SNPs as instruments for antihypertensive drug targets, I estimated their effect on systolic blood pressure using two-sample Mendelian randomization. The SNP-expression association, extracted from GTEx as described above, was on the scale of a standard deviation change in RNA expression levels for each additional effect allele. The SNP-systolic blood pressure association was extracted from the systolic blood pressure GWAS in UK Biobank and represented the standard deviation change in systolic blood pressure for each additional effect allele. Unlike published systolic blood pressure GWASs, this GWAS is not adjusted for body mass index to avoid bias in the two-sample Mendelian randomization estimates, which can potentially occur when both samples are not adjusted for the same factors. These associations were then used to estimate the effect of the protein target on systolic blood pressure (i.e. the standard deviation change in systolic blood pressure per standard deviation change in RNA expression levels). SNPs with evidence of an effect on systolic blood pressure were retained for the main analysis. This instrument selection process is presented in Figure 8.1.

8.2.5. Statistical methods

I used two-sample Mendelian randomization to estimate the effect of lowering systolic blood pressure on Alzheimer's disease in three ways. First, I estimated the effect of specific drug classes by combining the effects of any of the drug targets associated with a given drug class. This used the instruments defined in the previous section and summarized in Figure 8.1. Second, I estimated the effect of antihypertensive drugs as a

Figure 8.1: Flow chart summarizing the instrument selection process and Mendelian randomization analyses for the assessment of the effects of antihypertensive drugs on dementia prevention using genetic data.



MR: Mendelian randomization

whole on Alzheimer's disease by combining all drug targets. Again, this used the instruments defined in the previous section and summarized in Figure 8.1. Finally, I estimated the overall effect of systolic blood pressure on Alzheimer's disease by combining the effects of any genome-wide significant SNPs for systolic blood pressure. When multiple SNPs were being used as an instrument, 'clumping' was performed to identify independent SNPs using the linkage disequilibrium between them. SNPs absent in the outcome data were replaced by proxy SNPs in high linkage disequilibrium from the 1000 Genomes Project European data where possible. (104,213) Proxies were required to have a minimum R-squared value of 0.8 and palindromic SNPs were permitted if their minor allele frequency was less than 0.3.

Prior to the analysis, data were harmonised to represent an increase in systolic blood pressure. Mendelian randomization was then performed using the inverse variance weighted method or, for single-SNP instruments, the Wald ratio. (84,214,215) Once complete, the Mendelian randomization results were transformed to be the odds ratio for Alzheimer's disease per 10mmHg lower systolic blood pressure to make the effect comparable to taking an antihypertensive, which on average reduces systolic blood pressure by 9mmHg. (59) All analyses used genome reference consortium human build 37 (GRCh37), assembly Hg19, and were performed in R (version 3.4.4) using the package 'TwoSampleMR'. (104,195) All coding and editable results files are available from GitHub: https://github.com/venexia/MR-antihypertensives-AD.

8.2.6. Sensitivity analyses

As described in Section 4.6.2.3, Mendelian randomization estimates may be subject to horizontal pleiotropy. This is where the SNP(s) chosen to proxy the exposure affect the outcome by a different mechanism to that intended. (216) This invalidates the exclusion restriction assumption, which states that the instrument must only affect the outcome through the exposure (IV2 – Section 3.2.1). To estimate the extent of horizontal pleiotropy, I applied MR-Egger regression to all estimates based on ten or more SNPs. Note that a minimum number of SNPs is necessary to minimise weak instrument bias, which can affect MR-Egger analyses. The regression intercept for these analyses "can be interpreted as an estimate of the average pleiotropic effect across the genetic variants". (137) This can detect directional pleiotropy, which occurs when the biasing effects are

not balanced around the null. MR-Egger relies on the Instrument Strength Independent of Direct Effect (InSIDE) assumption, which specifies that the magnitudes of the genetic-variant association and pleiotropic effects should not be correlated. It is often referred to as "a weaker version of the exclusion restriction assumption". (137)

To examine heterogeneity within the drug classes, I also considered the effects of individual drug targets on Alzheimer's disease. This analysis allowed me to interrogate whether certain targets were driving the observed drug class effects. Drug classes with very heterogeneous target results can be considered to have less reliable estimates than those where targets were more homogeneous.

8.3. Results

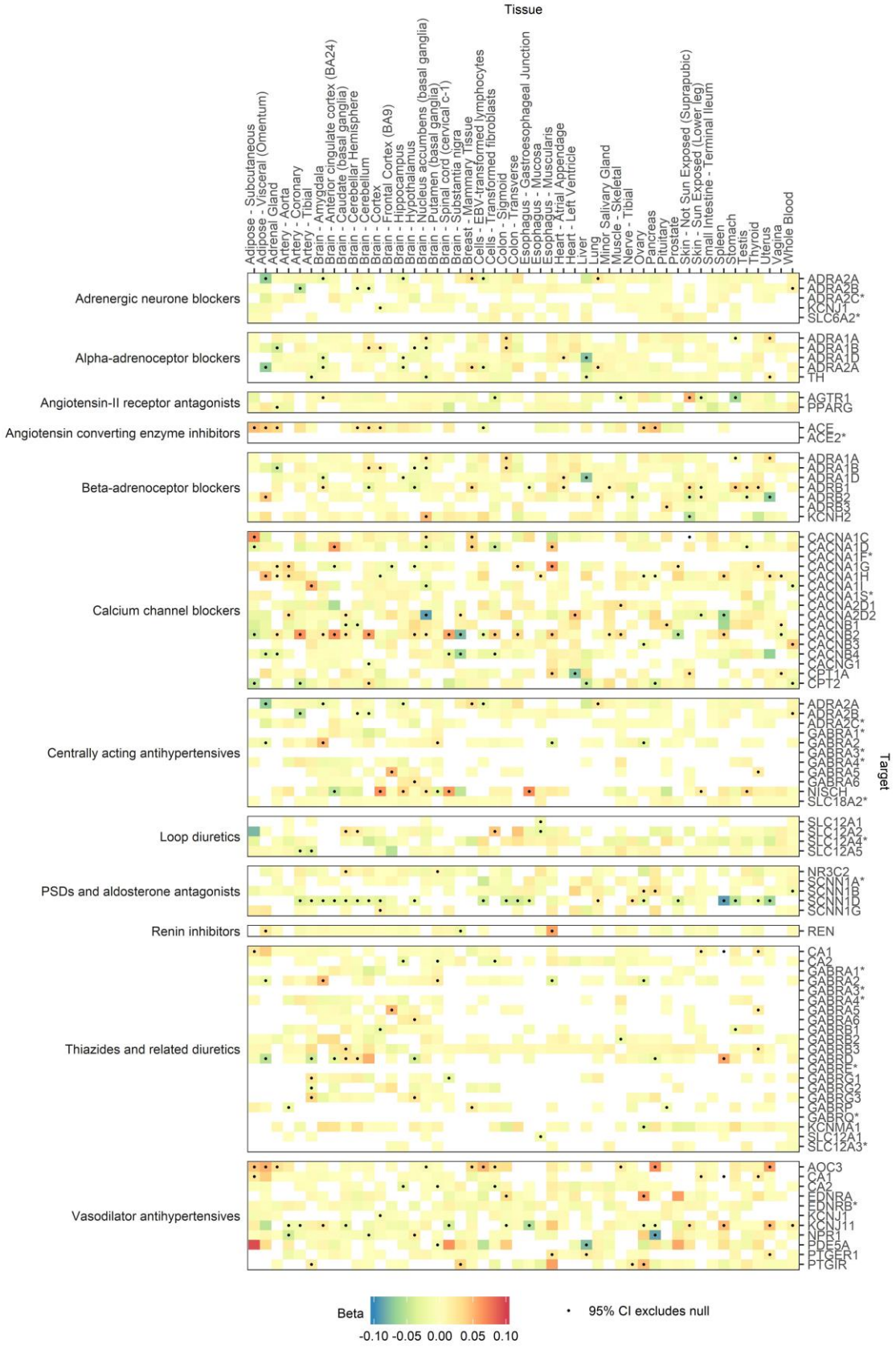
8.3.1. Instrument selection

I identified a total of 73 unique protein targets of antihypertensive drugs. Among these targets, 68 had an effect in one or more GTEx tissues and 58 of those 68 provided evidence that the target affected systolic blood pressure. Figure 8.2 summarizes the results of the Mendelian randomization analysis of expression on systolic blood pressure. A further six targets were excluded prior to the main analysis because neither the genetic instrument, nor a suitable proxy, were available in the outcome GWAS. Consequently, 52 unique protein targets were ultimately analysed.

8.3.2. Drug class effects

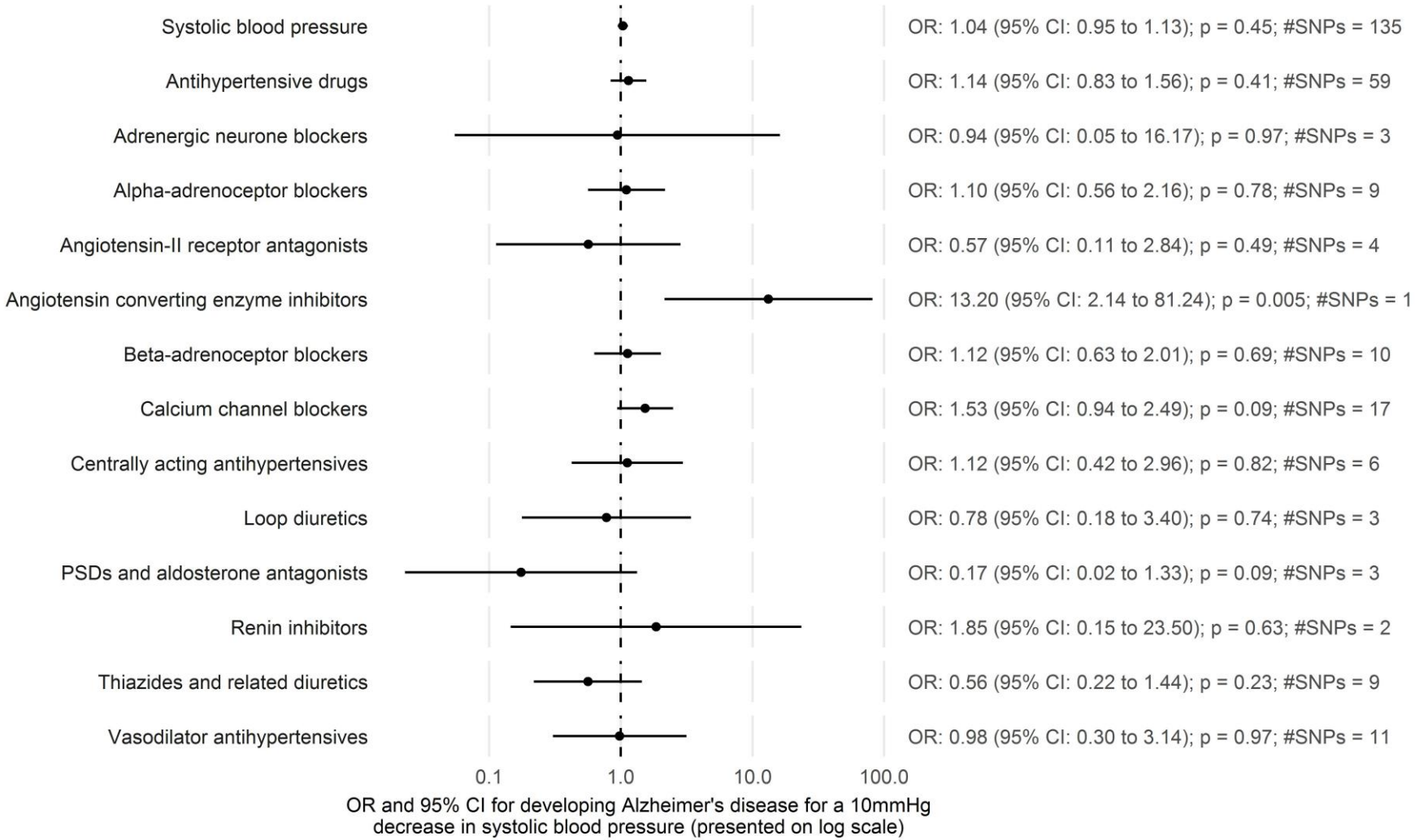
There was limited evidence that reducing systolic blood pressure affected risk of Alzheimer's disease at the drug class level with most estimates failing to exclude the null (Figure 8.3). For example, calcium channel blockers had an OR of 1.53 (95% CI: 0.94 to 2.49; p-value=0.09; SNPs=17) and loop diuretics an OR of 0.78 (95% CI: 0.18 to 3.40; p-value=0.74; SNPs=3) per 10mmHg lower systolic blood pressure. The exceptions to this were angiotensin-converting enzyme inhibitors (OR per 10mmHg lower systolic blood pressure: 13.20; 95% CI: 2.14 to 81.24; p-value=5.0×10^{-3}; rs4968783) and potassium-sparing diuretics and aldosterone antagonists (OR per 10mmHg lower systolic blood pressure: 0.17; 95% CI: 0.02 to 1.33; p-value=0.09; SNPs=3).

Figure 8.2: Heat map of estimates for the effect of gene expression on systolic blood pressure.



** No evidence for an effect on systolic blood pressure.*

Figure 8.3: Estimates for the effect of systolic blood pressure on Alzheimer's disease from two-sample Mendelian randomization.



8.3.3. Antihypertensive drug effect

There was little evidence for an overall effect of lowering systolic blood pressure on Alzheimer's disease when combining all identified drug targets (OR per 10 mmHg lower systolic blood pressure: 1.14; 95% CI: 0.83 to 1.56; p-value=0.41; SNPs=59) (Figure 8.3).

8.3.4. Systolic blood pressure effect

There was also little evidence for an overall effect of lowering systolic blood pressure on Alzheimer's disease, without consideration of the associated drugs, as indicated by the OR of 1.04 (95% CI: 0.95 to 1.13; p-value=0.45; SNPs=135) per 10 mmHg lower systolic blood pressure (Figure 8.3).

8.3.5. Sensitivity analyses

The Egger intercepts were close to zero for almost all analyses where they could be calculated (Table 8.1). In addition, the estimates from the inverse variance weighted and MR-Egger methods were similar for all analyses with both the point estimate and confidence interval for the inverse variance weighted method almost contained within the confidence interval for the MR-Egger method (Figure 8.4). Note that MR-Egger is susceptible to a lack of power hence the wide confidence intervals observed here. (137)

The analysis of individual targets identified some targets that were likely to be driving the drug class effects (Figure 8.5). For example, the target NR3C2 is estimated to have an OR per 10 mmHg lower systolic blood pressure of 2.01e-3 (95% CI: 5.22e-6 to 0.78; p-value=0.04; rs71616586) and is likely to have contributed to the extremely protective effect observed for potassium-sparing diuretics and aldosterone antagonists (OR per 10 mmHg lower systolic blood pressure: 0.17; 95% CI: 0.02 to 1.33; p-value=0.09; SNPs=3).

Table 8.1: Egger intercepts for the assessment of the effects of antihypertensive drugs on dementia prevention using genetic data.

Analysis	Intercept	Standard Error	P-value	SNPs
Overall	0.007	0.006	0.23	135
Combined	0.013	0.008	0.10	59
Beta-adrenoceptor blockers	0.001	0.018	0.95	10
Calcium channel blockers	0.018	0.013	0.18	17
Vasodilator antihypertensives	0.035	0.043	0.43	11

Figure 8.4: MR-Egger estimates for the effect of systolic blood pressure on Alzheimer's disease from two-sample Mendelian randomization.

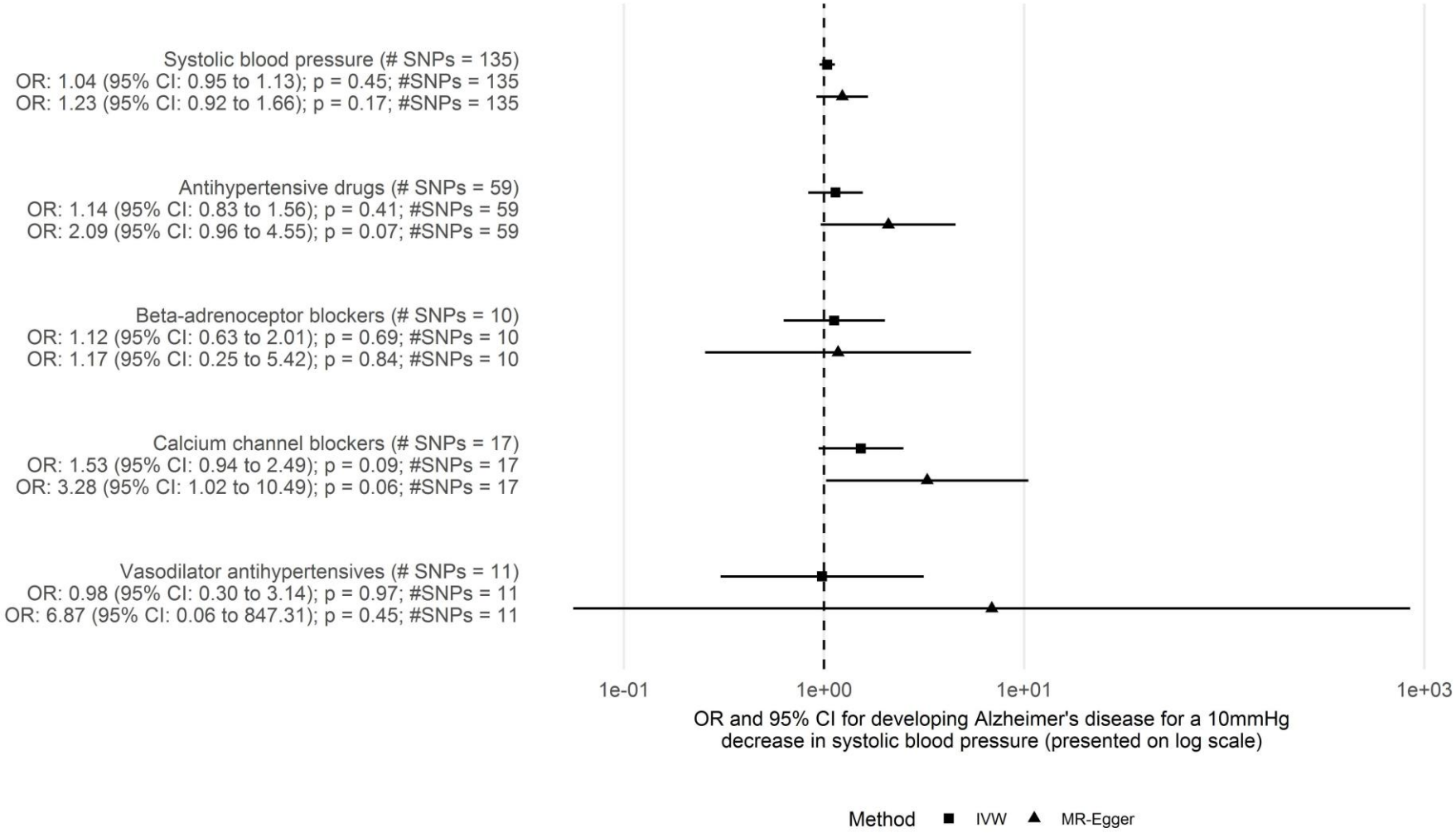
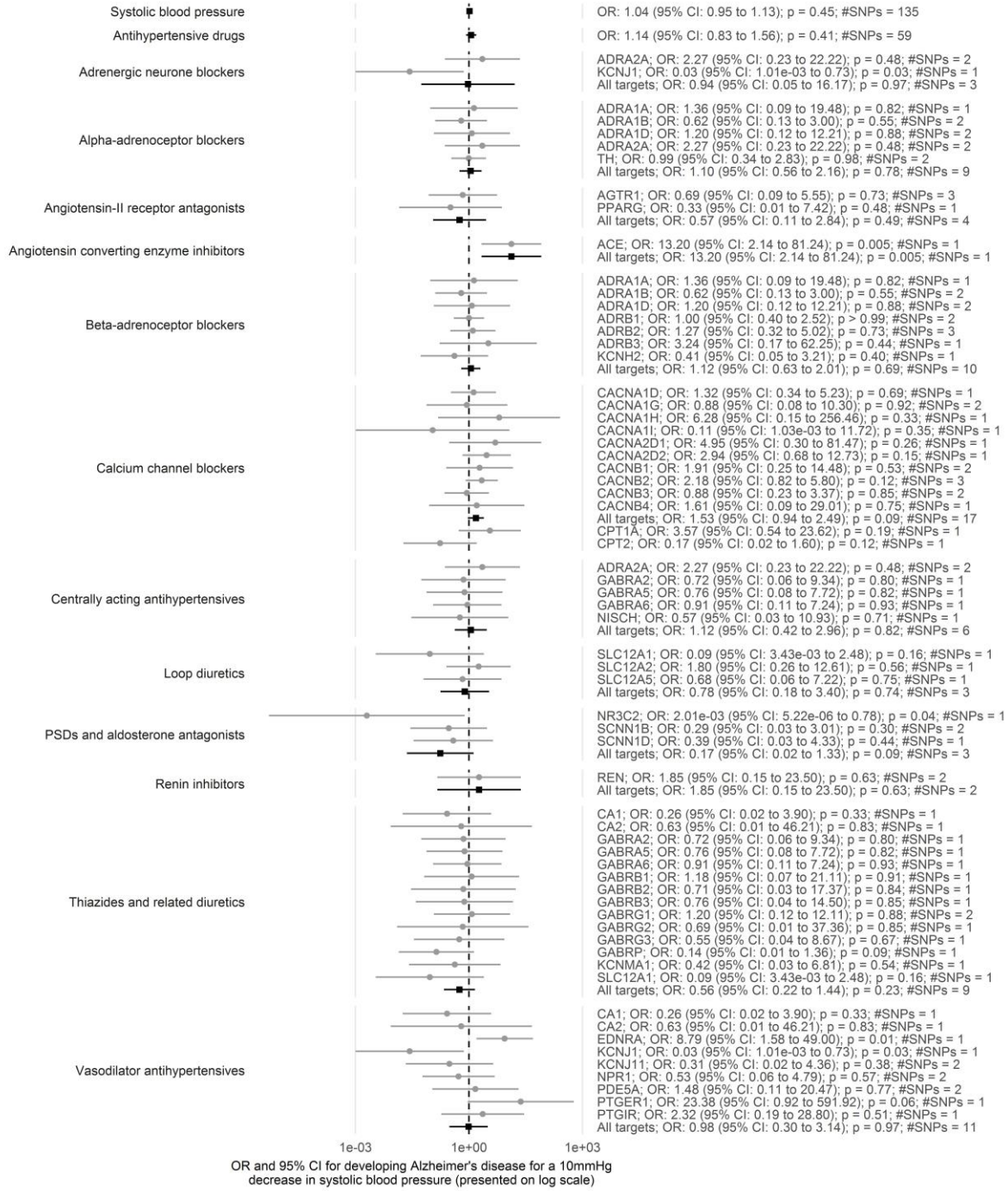


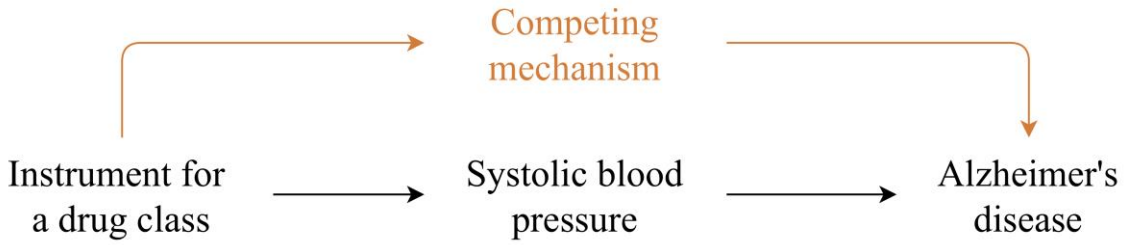
Figure 8.5: Target level estimates for the effect of systolic blood pressure on Alzheimer's disease from two-sample Mendelian randomization.



8.4. Discussion

There was limited evidence to support an overall effect of lowering systolic blood pressure on Alzheimer's disease risk (OR per 10 mmHg lower systolic blood pressure: 1.04; 95% CI: 0.95 to 1.13; p-value=0.45; SNPs=135). There was also limited evidence that lowering systolic blood pressure via antihypertensive drug classes affected Alzheimer's disease. For example, calcium channel blockers had an OR of 1.53 (95% CI: 0.94 to 2.49; p-value=0.09; SNPs=17) and vasodilator antihypertensives had an OR of 0.98 (95% CI: 0.30 to 3.14; p-value=0.97; SNPs=11) per 10mmHg lower systolic blood pressure. This was reflected in the overall effect of lowering systolic blood pressure on Alzheimer's disease when combining all identified drug targets, which had an OR of 1.14 (95% CI: 0.83 to 1.56; p-value=0.41; SNPs=59) per 10 mmHg lower systolic blood pressure. Despite this, there are some extreme results, such as angiotensin-converting enzyme inhibitors, which were associated with an increased Alzheimer's disease risk (OR per 10 mmHg lower systolic blood pressure: 13.29; 95% CI: 2.14 to 81.24; p-value=5.0×10^{-3}; rs4968783).

Figure 8.6: Illustration of Mendelian randomization in the presence of a competing mechanism.



The cause of these extreme results could be due to a competing mechanism, as illustrated in Figure 8.6. I estimated the effect of exposure to a given drug class on Alzheimer's disease using the effect of the instrument for that drug class on both systolic blood pressure (instrument-exposure association) and Alzheimer's disease (instrument-outcome association). The analysis assumed that the effect I was estimating acted through systolic blood pressure, however there is potentially a competing mechanism by which the given drug class can affect Alzheimer's disease. If a competing mechanism does exist and the instrument-exposure association (i.e. the effect of the drug class instrument on systolic

blood pressure) is small, estimates from Mendelian randomization can become inflated as the competing mechanism means the instrument-outcome association (i.e. the effect of the drug class instrument on Alzheimer's disease) remains large. This is more apparent if you consider the Wald ratio used to calculate the effect for single SNP instruments:

$$\text{Exposure-outcome association} = \frac{\text{Instrument-outcome association}}{\text{Instrument-exposure association}}$$

In this analysis, I found a small effect of systolic blood pressure on Alzheimer's disease and the extreme results were for drug classes that may well act through competing mechanisms. For instance, returning to the example of angiotensin-converting enzyme inhibitors, angiotensin-converting enzyme is proposed to affect vascular pathways (such as blood pressure), but also be responsible for the production of angiotensin-II that is thought to be associated with a number of the pathological processes (e.g. inflammation, oxidative stress, reduced blood flow) involved in Alzheimer's disease. (217) Importantly it is also thought have independent effects on amyloid beta. (9) In addition, potassium-sparing diuretics and aldosterone antagonists, which were also estimated to have an extreme effect (OR per 10 mmHg lower systolic blood pressure: 0.17; 95% CI: 0.02 to 1.33; p-value=0.09; SNPs=3), have previously been suggested to have a role, independent of blood pressure, in preventing cognitive decline.(218) This explanation for the extreme results observed for certain drug classes, along with the limited evidence for an effect among the remaining drug classes, indicates that antihypertensives are unlikely to have an effect on Alzheimer's disease via lowering systolic blood pressure.

Two previous Mendelian randomization studies have studied the overall effect of systolic blood pressure on Alzheimer's disease to date. These studies used different instruments and different systolic blood pressure GWAS, both to the present study and each other. (204,205) Østergaard et al found higher systolic blood pressure to be associated with a reduced risk of Alzheimer's disease, while Larsson et al found little evidence of an effect of systolic or diastolic blood pressure on risk of Alzheimer's disease. My results agree with Larsson et al in that there is unlikely to be an overall effect of systolic blood pressure on risk of Alzheimer's disease. Using the instruments reported by Østergaard et al and Larsson et al with the data used in the current study, I was able to reproduce their results

(Figure 8.7). There was a small overlap in the choice of SNPs used to instrument systolic blood pressure between my study and those from the literature (Østergaard et al: 14 SNPs; Larsson et al: 22 SNPs).

Gill et al recently conducted a study that combined Mendelian randomization using genetic variants related to antihypertensive targets with a PheWAS conducted in UK Biobank, however their analysis was restricted to beta-adrenoceptor blockers and calcium channel blockers. (206) Using the instruments reported by Gill et al with the data used in the current study, I was also able to reproduce their results (Figure 8.8). My results broadly agree with those reported by Gill et al for Alzheimer's disease, despite the minimal overlap in the choice of SNPs used to instrument beta-adrenoceptor blockers and calcium channel blockers (Beta-adrenoceptor blockers: 6 SNPs; Calcium channel blockers: 10 SNPs).

As discussed in Section 2.7, Larsson et al recently conducted a systematic review and meta-analysis on this topic. This study identified five randomized controlled trials that have investigated whether antihypertensives prevent dementia (not Alzheimer's disease specifically) with an overall relative risk of 0.84 (95% CI: 0.69 to 1.02; p-value=0.10). (12) It is worth reiterating that most studies described in the meta-analysis were from populations with high cardiovascular morbidity and were designed around cardiovascular related primary outcomes. This means that the proportion of dementia cases that derived from vascular mechanisms in these trials might be disproportionately high compared with other study populations. (60,61) This difference might explain the more favourable point estimate obtained in the meta-analysis compared with the results presented in this chapter.

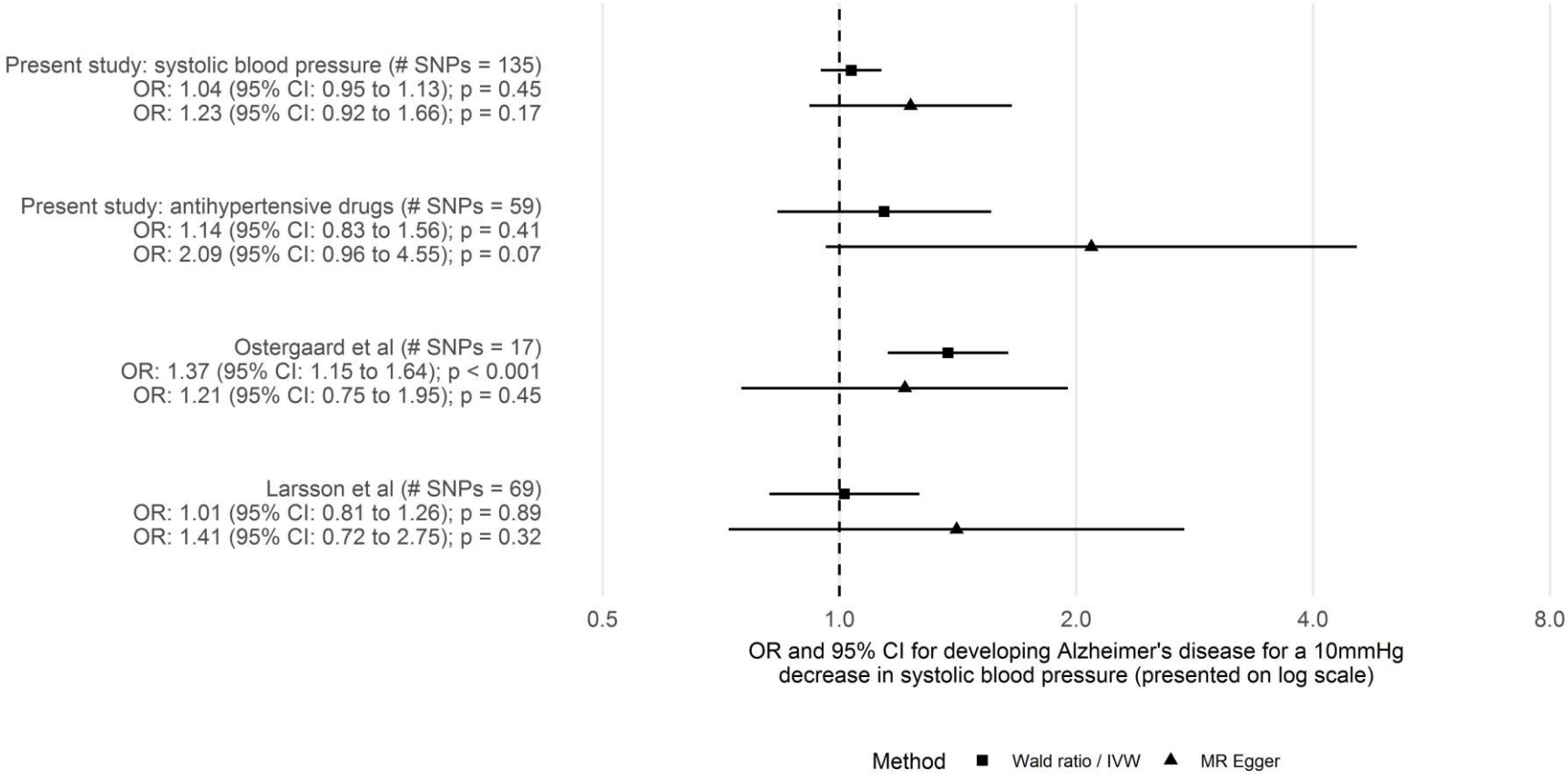
The SPRINT-MIND trial, which was published after the meta-analysis, and is also discussed in Section 2.7, assessed the effect of intensive vs standard blood pressure control using a range of antihypertensive medications on probable dementia, mild cognitive impairment and a composite outcome combining probable dementia and mild cognitive impairment. The trial found evidence to suggest that intensive blood pressure control was beneficial for the mild cognitive impairment and composite outcomes. Meanwhile, the estimate for the primary cognitive outcome of probable dementia

included the null within its confidence interval but may have been underpowered due to the early termination of the trial. There are several key differences between this trial and the analysis I present. First, my outcome of interest was Alzheimer's disease and so cannot be directly compared against the mild cognitive impairment or dementia outcomes used in the trial. Second, the trial was designed to compare treatment goals, whereas my analysis was comparing treatment with no treatment. Depending on blood pressure at baseline, these might yield different results. Third, as noted for the trials included in the meta-analysis conducted by Larsson et al, the primary outcome for the SPRINT trial was cardiovascular, meaning participants in SPRINT-MIND are more likely to have cognitive impairment and dementia outcomes derived from vascular mechanisms. Finally, Mendelian randomization estimates are of lifelong exposure, whereas this trial intervened on blood pressure for a median of 3.34 years in people with a mean age of 67.9 years. It is therefore possible that the trial has identified a critical period in which altering blood pressure has a beneficial impact on cognitive outcomes, which I cannot distinguish. Overall, while not directly comparable to our study, the findings from this trial, as well as the Larsson et al meta-analysis, provide further evidence concerning the repurposing of antihypertensives for Alzheimer's disease that can be considered together in a triangulation framework.

8.5. Strengths and limitations

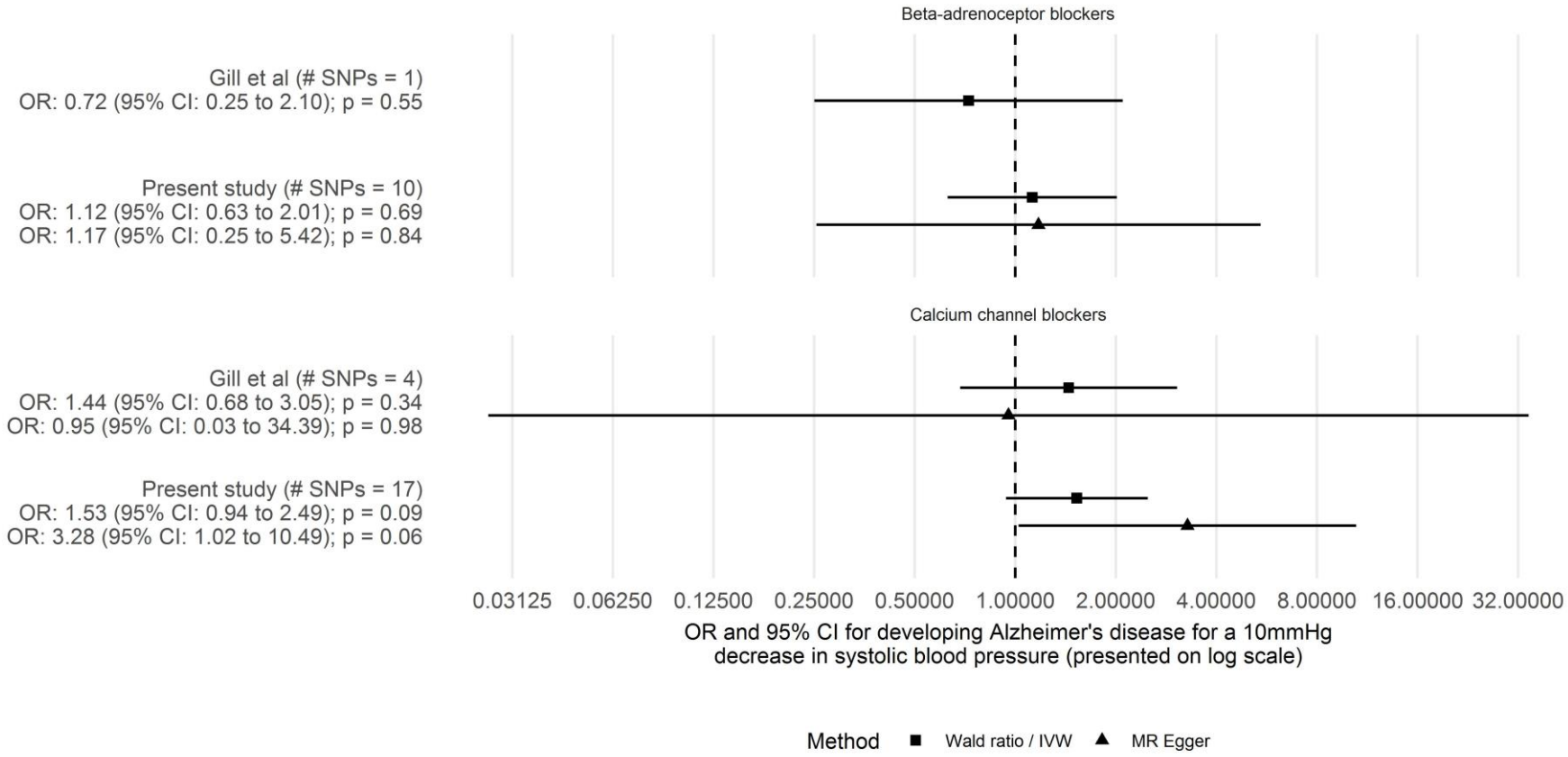
A strength of this study was the use of two-sample Mendelian randomization that meant I was able to use the International Genomics of Alzheimer's Project GWAS for the outcome data, which contains information on 17,008 Alzheimer's disease cases and 37,154 controls. (208) The use of Mendelian randomization, over more conventional pharmacoepidemiological approaches, will have also addressed certain forms of confounding. This includes confounding by indication and confounding by the environmental and lifestyle factors of patients, which cannot be fully adjusted for using observational data. This is because measurement error and incomplete capture of all these potential confounding factors inevitably leads to residual confounding. Further strengths of using Mendelian randomization to predict drug repurposing opportunities are provided in Section 4.6.1.

Figure 8.7: Estimates for the effect of systolic blood pressure on Alzheimer's disease from two-sample Mendelian randomization using previously reported systolic blood pressure instruments.



Note that this analysis implemented my pipeline, including clumping, so there are a reduced number of SNPs in the instruments for Ostergaard et al (originally 25) and Larsson et al (originally 93) as the clumping criteria are likely to differ to those previously used.

Figure 8.8: Estimates for the effect of systolic blood pressure on Alzheimer's disease from two-sample Mendelian randomization using previously reported antihypertensive drug instruments.



Note that this analysis implemented my pipeline, including clumping, so there are a reduced number of SNPs in the instruments for beta-adrenoceptor blockers (originally 6) and calcium channel blockers (originally 24) obtained from Gill et al as the clumping criteria are likely to differ to those previously used.

The limitations of this study are discussed in general in Section 4.6.2. They include the risk of horizontal pleiotropy, which I addressed in this study by conducting sensitivity analyses using MR-Egger when possible. Sensitivity analyses that considered the individual drug target effects also identified some heterogeneity that may have affected the drug class estimates. For example, the estimate for potassium-sparing diuretics and aldosterone antagonists may have seemed more protective due to the particularly large protective effect observed for one of the three targets under consideration: NR3C2. I was also limited in this study by the fact that Mendelian randomization estimates the effect of lifelong exposure, while drugs typically have much shorter periods of exposure. This means that the estimated effect sizes will not directly reflect what is observed in trials or clinical practice and may not distinguish critical periods of exposure. (138)

8.6. Summary

This study provided further new evidence about the potential effects of antihypertensives on risk of dementia by using a different method, subject to different biases, to assess this research question. I found little evidence to suggest that lowering systolic blood pressure itself will affect risk of developing Alzheimer's disease. This was accompanied by limited evidence for many of the antihypertensive drug classes that I tested. Despite this, there was some suggestion of an effect through non-blood pressure lowering mechanisms for some drug classes, namely: angiotensin-converting enzyme inhibitors and potassium-sparing diuretics and aldosterone antagonists. This study and the non-genetic instrumental variable analysis study discussed in Chapter 7 will be considered together in Chapter 9 as two forms of evidence concerning the potential repurposing of antihypertensives for Alzheimer's disease prevention.

Chapter 9. Discussion

9.1. Summary

In this chapter, I summarize the outputs relating to the specific aims of this thesis. First, I describe the development of a power calculator for non-genetic instrumental variable analysis studies in the context of pharmacoepidemiology. Second, I cover how Mendelian randomization can be used to predict drug repurposing opportunities. Third, I summarise my examination of the impact of regulatory guidance and patent expiry on dementia drug prescribing. Fourth, I describe the assessment of antihypertensive drugs using instrumental variable analysis with electronic health record data and, finally, I describe the assessment of antihypertensive drugs using instrumental variable analysis with genetic data. This chapter also includes a summary of my efforts to make my research open and reproducible and details of potential future work arising from this thesis. The chapter concludes with a discussion that draws everything together including the novel findings, innovative methods, overarching limitations and future work associated with this thesis.

9.2. Power calculator for instrumental variable analysis in pharmacoepidemiology

9.2.1. Principal findings

In Chapter 3, I derived a power formula for studies using a single binary instrument to analyse the causal effect of a binary exposure on a continuous outcome in the context of pharmacoepidemiology. I then validated the formula using a simulation that generated realistic data where the power of the study was known. Finally, to allow others to easily implement the formula, I produced an online calculator and packages for R and Stata that are publicly available.

9.2.2. Comparison with existing literature

Prior to the development of my power calculator, the majority of power calculators for instrumental variable analysis had been developed for Mendelian randomization. (29,30) However, pharmacoepidemiological research questions have distinct structures that are

not sufficiently catered for by Mendelian randomization power calculators. This is because, unlike Mendelian randomization studies that often use a case-control study design, pharmacoepidemiology studies typically use a cohort study design. Further to this, pharmacoepidemiology studies usually report a risk difference for a binary exposure using a binary instrument, while Mendelian randomization studies report on a continuous exposure using a discrete or continuous genetic instrument (count of alleles or allele score respectively). As a result of these differences, as well as the stronger instruments and larger causal effects seen in pharmacoepidemiology, there was a need for a dedicated power calculator for instrumental variable analysis in the context of this field.

9.2.3. Strengths and limitations

The key strength of this power calculator is that it has been developed specifically for questions in the context of pharmacoepidemiology, which the existing power calculators did not cater for. It has also been designed to be accessible, with publicly available tools developed to allow others to implement the formula. However, the limitations of this research are that the power formula is only applicable when conducting an instrumental variable analysis study that has a single binary instrument, binary exposure and continuous outcome. It is also limited, like every power formula, by its parameters as they simplify the data and so can miss underlying complexities.

9.2.4. Implications of this research

The main implication of this research is the ability for others to calculate the power of instrumental variable studies using a single binary instrument, binary exposure and continuous outcome in the context of pharmacoepidemiology with ease, via the online power calculator and packages in R and Stata. (31) As noted in Section 3.4.1, this is increasingly important for grant and data request applications and as the interest in more complex analytical methods, such as instrumental variable analysis, in pharmacoepidemiology increases.

9.3. Mendelian randomization for predicting drug repurposing opportunities

9.3.1. Principal findings

In Chapter 4, I introduced Mendelian randomization as a novel approach for predicting drug repurposing opportunities. Specifically, I demonstrated the use of Mendelian randomization for predicting drug repurposing opportunities through examples from the existing literature and discussed the strengths and limitations of using this method for this purpose. Within the chapter, I also advocate the synthesis of evidence from Mendelian randomization and other approaches, in the spirit of triangulation, to improve causal inferences concerning drug effects. (20) I later provide an example of Mendelian randomization for drug repurposing through the application of this method to assess whether antihypertensives can be repurposed for dementia prevention in Chapter 8.

9.3.2. Comparison with existing literature

There is a large existing literature on Mendelian randomization that includes both methodological developments and applications of the method. (84,85,138,219) There is also a large existing literature concerning the use of genetics in drug discovery. (87,220–222) However, Mendelian randomization has not been used to its full potential for predicting drug repurposing opportunities to date and discussion of specific pharmacoepidemiology issues, such as confounding by indication, is uncommon in the Mendelian randomization literature. This chapter and the accompanying article therefore highlight the potential of Mendelian randomization for this purpose and discuss the strengths and limitations that are specific to this application. (32)

9.3.3. Strengths and limitations

There are many strengths and limitations associated with the use of Mendelian randomization for predicting drug repurposing opportunities. Strengths include that it addresses confounding by environmental and lifestyle factors, confounding by indication, and reverse causation. All of which are of concern when conducting pharmacoepidemiology studies and may be particularly problematic for later life conditions such as dementia. Limitations include the risk of horizontal pleiotropy, weak

instrument bias, and the fact that estimates reflect lifelong exposure, which may not be realistic in terms of clinical intervention and may obscure critical periods of exposure.

9.3.4. Implications of this research

This research has contributed to the discussion of using Mendelian randomization for drug discovery in the literature, a key area of active research. The process of writing this chapter and the accompanying article has allowed me to collect together several examples of Mendelian randomization being used to predict drug repurposing opportunities. This has proved a useful tool for teaching – for example, the issues discussed in this chapter formed the basis of a lecture entitled 'Use of Mendelian randomization in drug discovery and target validation' for the Bristol Medical School Mendelian randomization short course in 2018 and 2019.

9.4. Factors effecting existing dementia drug prescribing

9.4.1. Principal findings

In Chapter 6, I examined prescribing trends in England from the launch of the drugs for dementia up to 1st January 2016, using data from the CPRD. I found that the overall trend was for increasing prescriptions in each drug class over the period in which they were studied. I also found that prescriptions of acetylcholinesterase inhibitors increased at the end of 2012, probably in response to the patent expiry of these drugs earlier that year and, potentially, the Prime Minister's Dementia Challenge launched in May 2012. However, neither this strategy nor patent expiry appeared to influence prescriptions of NMDA receptor antagonists. Instead trend changes in this drug class were driven by NICE guidance released in 2011 that allowed access to these drugs outside of clinical trials.

9.4.2. Comparison with existing literature

The existing literature concerning factors influencing dementia drug prescribing in England has mainly focused on the impact of the National Dementia Strategy. (33,34) My findings were broadly in agreement with the existing literature in that there has been an increasing trend in prescribing of dementia drugs over time. However, I identified a trend change in 2012 as the largest factor influencing prescriptions rather than the launch of the National Dementia Strategy in 2009 that they consider. The main difference between my study and the existing literature is the use of a hypothesis-free approach that allowed me to consider multiple factors simultaneously.

9.4.3. Strengths and limitations

The key strengths of this research are the large sample of primary care data and the long-follow up of patients within this sample. The large sample is important to maintain statistical power for the analysis, which involves multiple tests, while the long follow-up allowed patients who experienced a treatment delay to remain in the analysis. The limitations of this research include the potential for misclassified outcomes that affects all studies using electronic health record data as they are not designed for research use. This research may also be affected by difficulties in determining the time lag between events that may have influenced prescribing and the observed trend changes in prescribing, which make it difficult to determine causality.

9.4.4. Implications of this research

This research has identified several factors that may have affected prescriptions of dementia drugs over a twenty-year period. (35) I found that dementia drug prescribing does not always respond to factors such as regulatory guidance, recommendations, or patent expiry, and when it does, not necessarily in a predictable way. This has implications in terms of how changes to regulatory guidance and recommendations are communicated to clinicians in the future and how new drugs, including repurposed candidates, are used. In addition, this study provides knowledge of the dementia drug prescribing trends, which is necessary when adjusting for these medications in other analyses, such as studies of progression, and when assessing their effectiveness.

9.5. Assessment of antihypertensive drugs using electronic health record data

9.5.1. Principal findings

In Chapter 7, I conducted an instrumental variable analysis using physicians' prescribing preference as an instrument in electronic health record data. I found small differences in drug class effects on risk of probable Alzheimer's disease, possible Alzheimer's disease, vascular dementia, other dementias and any dementia. However, I showed the magnitude of the differences between drug classes is smaller than many observational studies have previously reported for these outcomes. I also showed that the bias in this study is likely to be less, or at least no more extreme, than that observed for multivariable linear regression.

9.5.2. Comparison with existing literature

The evidence for repurposing antihypertensives for dementia prevention from studies with an active comparator, such as that presented in Chapter 7, is summarized in Table 9.1. The key difference between my analysis approach and those previously used is the assumptions that they rely on. For the instrumental variable analysis, I must assume relevance, the exclusion restriction, independence and monotonicity, whereas the non-instrumental variable analysis observational studies assume no residual confounding. Violation of these assumptions may explain the difference in magnitude observed. Neither of these sets of assumptions can be tested fully however, there was an indication of potential residual confounding when I compared the instrumental variable analysis with an equivalent multivariable linear regression analysis. Despite this, the analyses broadly agree in that there is likely to exist some differential effects among antihypertensives classes in terms of dementia prevention.

9.5.3. Strengths and limitations

Like the examination of dementia drug prescribing trends, this study benefits from a large cohort of patients from the CPRD. For this study, this is particularly beneficial as it would be infeasible to obtain such a large sample in the equivalent randomized controlled trial. A further strength is the use of instrumental variable analysis, which

contrasts with the other methods that have been implemented prior to this study that have all assumed no residual confounding. Instead, this study relies on alternative assumptions and so is subject to different biases. This allows triangulation of evidence. (20) The key limitation of this study is that, other than the assumption of relevance, I am not able to test the instrument assumptions. The analysis also remains susceptible to misclassified outcomes as it uses electronic health record data that are not intended for research.

9.5.4. Implications of this research

This research provides a new source of evidence concerning the potential repurposing of antihypertensives for the prevention of dementia. (21) As noted at several points in this thesis, this is important as the current evidence concerning this hypothesis is inconclusive. This study is also one of the largest to investigate the potential repurposing of antihypertensives to date with 849,378 patients exposed to the drugs of interest and total follow-up of 5,497,266 patient-years. It is also one of the first studies (along with the Mendelian randomization presented in Chapter 8) to use causal inference methods to test this hypothesis and should be more robust to residual confounding than previously used study designs.

9.6. Assessment of antihypertensive drugs using genetic data

9.6.1. Principal findings

In Chapter 8, I conducted a Mendelian randomization analysis, which found limited evidence that lowering systolic blood pressure, via antihypertensive drug classes, affected Alzheimer's disease risk. I also found limited evidence for an effect of lowering systolic blood pressure on Alzheimer's disease when combining all drug targets and without consideration of the associated drug targets. Considered together, this would suggest that if specific antihypertensive drug classes do affect Alzheimer's disease risk, they are unlikely to do so via systolic blood pressure.

Table 9.1: Summary of evidence for repurposing antihypertensives for dementia prevention from studies using an active comparator.

Analysis	Outcome	Summary of results	Assumptions	Potential sources of bias
Instrumental variable analysis using electronic health record data presented in Chapter 7	Probable Alzheimer's disease; possible Alzheimer's disease; vascular dementia; other dementias; any dementia	Small differences in drug class effects on risk of dementia outcomes	Relevance; exclusion restriction; independence; monotonicity	Cannot prove all instrument assumptions are met; exposure misclassification; outcome misclassification; weak instrument bias
Meta-analysis by Larsson et al of seven observational studies	Dementia or Alzheimer's disease	One study (of three) suggesting angiotensin-converting enzyme inhibitors were protective (11–13); three studies (of four) suggesting angiotensin-II receptor blockers were protective (13–16); and one study (of one) suggesting calcium channel blockers were protective (17)	No residual confounding	Unmeasured confounding; exposure misclassification; outcome misclassification; missing data
Retrospective cohort study by Barthold et al comparing renin-angiotensin system acting and non-renin-angiotensin system acting antihypertensive drugs	Alzheimer's disease	Angiotensin-II receptor blockers may reduce the risk of Alzheimer's disease in certain groups, namely white and black women and white men	No residual confounding	Unmeasured confounding; exposure misclassification; outcome misclassification; missing data

9.6.2. Comparison with existing literature

The evidence for repurposing antihypertensives for dementia prevention from studies with a non-active comparator, such as that presented in Chapter 8, is summarized in Table 9.2. My results were in line with those from the Larsson et al meta-analyses of clinical trials and observational studies, which found weak evidence to support repurposing antihypertensive drugs for dementia prevention. Given that the instrumental variable analysis using genetic data relies on different assumptions to the previously conducted analyses and is subject to different biases, this furthers the case against repurposing antihypertensives for dementia prevention via blood pressure lowering mechanisms. However, it does not exclude other non-blood pressure lowering mechanisms of antihypertensives – such as those proposed for renin-angiotensin system acting drugs as a result of the angiotensin hypothesis of Alzheimer's disease. This hypothesis proposes that additional Alzheimer's disease-related pathological processes, independent of any blood pressure effects, such as cognition, inflammatory and oxidative stress mechanisms – and potentially activities related to amyloid beta and tau pathology – are associated with renin-angiotensin system dysfunction. (217)

9.6.3. Strengths and limitations

The use of Mendelian randomization for this study will have addressed concerns regarding confounding by indication, environmental or lifestyle factors, and reverse causation, that may have been present for the more conventional pharmacoepidemiology studies. Specifically, the use of two-sample Mendelian randomization (compared to one-sample Mendelian randomization) will have increased the power of this study as it allowed the use of a large GWAS of Alzheimer's disease, maximizing the number of cases in the analysis. The main limitation of this study, as with all Mendelian randomization studies, is the risk of horizontal pleiotropy that occurs when the SNPs influence the outcome through a pathway other than that of interest. The inference from this study is also limited by the fact that the estimates refer to lifelong exposure, however drug exposure is likely to be shorter and occur later in life. This means estimates will not directly reflect those obtained from other study designs or observed in clinical practice.

Table 9.2: Summary of evidence for repurposing antihypertensives for dementia prevention from studies using a non-active comparator.

Analysis	Outcome	Summary of results	Assumptions	Potential sources of bias
Instrumental variable analysis using genetic data presented in Chapter 8	Alzheimer's disease	Limited evidence that lowering systolic blood pressure via antihypertensive drug classes affected Alzheimer's disease risk	Relevance; exclusion restriction; independence; monotonicity	Horizontal pleiotropy; weak instrument bias
Meta-analysis by Larsson et al of five clinical trials	Dementia	Four studies of five found a weak protective effect, but the meta-analysis failed to exclude the null	Trial arms are comparable with the exception of the treatment they receive	Loss to follow-up could differ in the treatment arms; most of the trials were from populations with high cardiovascular morbidity and were designed around cardiovascular related primary outcomes
Meta-analysis by Larsson et al of five observational studies	Alzheimer's disease	Weak evidence for an effect on Alzheimer's disease	No residual confounding	Unmeasured confounding; exposure misclassification; outcome misclassification; missing data
Meta-analysis by Larsson et al of three observational studies	Dementia	Weak evidence for an effect on dementia, which failed to exclude the null	No residual confounding	Unmeasured confounding; exposure misclassification; outcome misclassification; missing data
SPRINT-MIND clinical trial	Mild cognitive impairment and dementia	Intensive blood pressure control was beneficial for the mild cognitive impairment and composite outcomes compared with standard blood pressure control	Trial arms are comparable with the exception of the treatment they receive	Loss to follow-up could differ in the treatment arms

9.6.4. Implications of this research

Like the application of instrumental variable analysis using electronic health record data, the application of instrumental variable analysis using genetic data also provides new evidence concerning repurposing antihypertensives for dementia prevention and is one of the first studies to use causal inference methods. (22) This new evidence is subject to different biases than the existing literature and uses a different data source. Consequently, it is a key component of evidence when triangulating the evidence concerning this hypothesis.

A further implication of this research is that is provides a 'real-world demonstration' of the use of Mendelian randomization for drug repurposing as I proposed in Chapter 4 and in the article 'Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities'. (32)

9.7. Open research

Throughout my PhD, I have endeavored to make my research as open and reproducible as possible. I have summarized these efforts in Table 9.3.

9.8. Future work

9.8.1. Power formulae for other types of instrumental variable analysis in the context of pharmacoepidemiology

A natural extension to the work described in this thesis would be to develop power formulae applicable for instrumental variable analyses other than those using a single binary instrument to analyse the causal effect of a binary exposure on a continuous outcome in the context of pharmacoepidemiology. In particular, there is a need to develop power formulae that can be applied when using a binary outcome, such as that studied in this thesis. The addition of formula for different binary/continuous instrument-exposure-outcome combinations to the existing online tool, and packages in R and Stata, would greatly enhance the research I conducted.

Table 9.3: Summary of the open research efforts associated with this thesis.

Chapter	Data and coding files	Paper
Chapter 3. Methods: Instrumental variable analysis using electronic health record data	The code for the simulations, the Shiny application and the packages in R and Stata is available from GitHub: https://github.com/venexia/PharmIV	Available open access from the International Journal of Epidemiology, previously uploaded to bioRxiv.
Chapter 4. Methods: Instrumental variable analysis using genetic data	Not applicable.	Available open access from the International Journal of Epidemiology, previously uploaded to bioRxiv.
Chapter 5. Methods: The Clinical Practice Research Datalink	Restrictions apply to the availability of CPRD data, which was used under license for this work. The code lists used to extract this data are publicly available: https://doi.org/10.5523/bris.1plm8il42rmlo2a2fqwslwckm2 The code used to clean the CPRD data is available from GitHub: https://github.com/venexia/CleanCPRD	Available open access from BMJ Open.
Chapter 6. Results: Prescribing trends for drugs for dementia	Restrictions apply to the availability of CPRD data, which was used under license for this work. The code lists used, in addition to those used in Chapter 5, are publicly available: https://doi.org/10.5523/bris.2h4rmk9v7pw2k23h7vgf9tx1ea The code used for this analysis is available from GitHub: https://github.com/venexia/DementiaDrugsCPRD	Available open access from Alzheimer's & Dementia, previously uploaded to PeerJ Preprints.
Chapter 7. Results: Assessment of the effects of antihypertensive drugs on dementia prevention using electronic health record data	Restrictions apply to the availability of CPRD data, which was used under license for this work. The code lists used, in addition to those used in Chapter 5, are publicly available: http://rebrand.ly/repurposing-antihypertensives-dementia-codelists. The code used for this analysis is available from GitHub: https://github.com/venexia/repurposing-antihypertensives-dementia	Available from bioRxiv. Currently under peer-review.
Chapter 8. Results: Assessment of the effects of antihypertensive drugs on dementia prevention using genetic data	The data used in this chapter are publicly available and have been obtained from the sources listed in the README for the following GitHub repository: https://github.com/venexia/MR-antihypertensives-AD, which also contains the code used for this analysis.	Available open access from the International Journal of Epidemiology, previously uploaded to bioRxiv.

9.8.2. MR-PheWAS using electronic health record data

In Chapter 4, I suggest that the application of MR-PheWAS to a combination of genetic data and electronic health records could be particularly powerful. At the time of writing, this was not possible. However, the continued development of resources, such as UK Biobank and the Million Veteran Program, throughout my PhD means that we will soon be able to implement this type of analysis. (223,224) Future work should look to combine the developments that have been made in instrument selection for drug proxies – such as that presented in Chapter 8 – with an MR-PheWAS using electronic health record data.

9.8.3. Assessment of antihypertensives for dementia progression

This thesis has focused on repurposing antihypertensives for dementia prevention but, given the suggested links between hypertension and dementia (Chapter 2), it is possible that these drugs may have utility for dementia progression. As mentioned in the protocol for the CPRD element of this thesis, this could be studied using existing electronic health record data. (36) However, further work would be needed to conduct a Mendelian randomization analysis due to the current lack of GWAS on dementia progression. As noted in Chapter 4, the lack of progression GWAS is not specific to the dementia field but is an issue that requires increased focus across disease areas.

9.9. Conclusions

There are three key novel findings, corresponding to each of the results chapters, that have arisen from this thesis. Firstly, prescriptions of dementia drugs have increased since their launch, but a join point analysis of prescribing trends suggests that the different classes of drugs have been affected by different factors during this time (Chapter 6). For instance, acetylcholinesterase inhibitor prescriptions responded most to expiry of their patents, while NMDA receptor antagonists responded most to changes in NICE guidance. Secondly, instrumental variable analysis using physician's prescribing preference suggests that small differences exist between antihypertensive drug classes in terms of their effect on dementia prevention (Chapter 7). However, the magnitude of the differences is smaller than many observational studies have previously reported. Finally, Mendelian randomization provides limited evidence that lowering systolic blood

pressure, via antihypertensive drug classes, affects Alzheimer's disease risk (Chapter 8). This suggests that if specific antihypertensive drug classes do affect the risk of Alzheimer's disease, they may not do so via systolic blood pressure.

An important aspect of these novel findings is the innovative methods that have been applied to obtain them. An example of this is the use of joinpoint analysis to study the factors affecting prescribing of dementia drugs (Chapter 6). This allowed a hypothesis-free approach to determine the key factors affecting trends rather than the targeted approaches used previously in the literature. A further example is the application of instrumental variable analysis to investigate whether antihypertensive drugs could be repurposed for dementia prevention (Chapters 7 and 8). Although both instrumental variable analysis using physician's prescribing preference and Mendelian randomization have been applied in many therapy areas before, they have not previously been used for this research question. This thesis has also contributed an innovative method to the literature through the derivation of a novel power formula for instrumental variable studies using a single binary instrument to analyse the causal effect of a binary exposure on a continuous outcome in the context of pharmacoepidemiology, as well as the development of packages in R and Stata to implement it (Chapter 3).

A key consideration when discussing the potential repurposing of antihypertensives for dementia prevention is biological plausibility. As such, plausibility is one of the nine Bradford Hill criteria outlined by Sir Austin Bradford Hill in 1965 for assessing causality. (225) As reasoned in Section 1.2, antihypertensive drugs have been proposed as drug repurposing candidates for dementia for several reasons that relate to their biological plausibility. These include the observed associations between midlife hypertension and later-life risk of Alzheimer's disease and vascular dementia (5–8); the increasing recognition that one of the earliest pathological events in the development of Alzheimer's disease is vascular dysregulation (9); and suggestions that some antihypertensives may have other neurological benefits. However, an important aspect of each of these reasons is the timing of the intervention. Alzheimer's disease is thought to have a long prodromal phase starting in midlife. Consequently, careful consideration must be given as to when an intervention would need to occur in order to modify the disease course. The instrumental variable analysis using electronic health record data that I present in this

thesis is limited in that I could not specify when in the life course antihypertensives were initiated. This meant the cohort for my analysis had a median age of 61 (interquartile range: 51 to 71) at index. This is likely to be too late in the life course for a primary prevention strategy (i.e. an intervention implemented prior to the onset of disease) to have an effect but may be informative when considering these drugs as a secondary prevention strategy (i.e. an intervention implemented after the onset of disease but before clinical symptoms). (51) The instrumental variable analysis using genetic data had a different relationship with timing because this analysis considered the lifelong effect of exposure to high blood pressure. A disadvantage of this approach is that it removes the opportunity to identify critical periods of exposure and means that exposures with mixed beneficial and adverse effects during the life course can appear to have a null overall effect. However, if the assumptions necessary for Mendelian randomization hold – specifically the "assumption that the proximal association of the genetic variant is with the risk factor, not with the outcome (nor with an alternative cause of the outcome)" – then the exposure must occur prior to the outcome. (226) This allows for assessment of the exposure as a primary prevention strategy, which can be difficult to assess using other study designs.

This thesis has several limitations. Those pertaining to individual studies have been discussed in previous sections in this chapter however, there are some common themes that apply across studies. Firstly, this thesis uses secondary data, such as that from the CPRD and GWAS summary statistics. These data were not collected specifically for the research questions studied. This has implications in terms of the types of analyses that could be applied and aspects of the analyses, such as diagnosis definitions and available covariates. Secondly, several of the analyses presented have had issues associated with time. In particular, the analyses concerning repurposing antihypertensives may be limited with regards to the timing and duration of exposure they can consider – as discussed above. A further example of a time related issue is the difficultly in determining the time lag between events that may have influenced prescribing and the observed trend changes in the CPRD for the dementia drug prescribing trend analyses in Chapter 6. Thirdly, dementia as defined in much of the data used in this thesis represents a range of symptoms, such as memory loss, and not necessarily degenerative brain disease. This may mean that the data captures several distinct conditions that share symptoms. If these

conditions differ in their causes, then causal effects estimated using the data may be attenuated due to the dilution of the outcome of interest. Furthermore, defining disease based on symptoms may also have implications regarding how the diagnosis is recorded – for example, those with more extensive vascular history may be misclassified as having vascular dementia. A final limitation of this thesis is specific to the analyses concerning repurposing antihypertensives for dementia prevention and is that these analyses were linked to the primary indication of the drugs, i.e. blood pressure lowering. In the assessment of antihypertensive drugs using electronic health record data, this occurred because patients were required to have hypertension (or another existing indication of these drugs) to receive the drugs. Consequently, the results obtained from these analyses may not generalize to other populations, such as those without hypertension. In the assessment of antihypertensive drugs using genetic data, this was the case because the instruments were chosen in such a way that the observed effects had to act through lowering systolic blood pressure. This meant I was unable to make inferences concerning drug effects that were independent of systolic blood pressure.

The novel aspects of this thesis could be used to inform future work. For instance, while the derivation of a novel power formula for instrumental variable studies using a single binary instrument to analyse the causal effect of a binary exposure on a continuous outcome in the context of pharmacoepidemiology is a novel addition to the literature (Chapter 3), there are further formulae that could be derived. As proposed in Section 9.8.1, a natural extension would be to consider formulae for different binary/continuous instrument-exposure-outcome combinations. The aim of this would be to ultimately provide a complete set of formulae for any instrumental variable analysis conducted in the context of pharmacoepidemiology. This thesis may also inform future work concerning the use of antihypertensive drugs for the treatment of dementia progression. One of the difficulties of studying disease progression is disentangling factors that are influencing incidence. Evidence concerning disease incidence, such as that presented in this thesis, can therefore be useful when interpreting results related to disease progression.

To conclude, this thesis has provided new evidence concerning the potential repurposing of antihypertensives for dementia prevention. Instrumental variable analysis using

electronic health record data suggested that, while differences between antihypertensives used for this purpose exist, they are smaller than previously reported in the literature. Meanwhile, instrumental variable analysis using genetic data suggested that there was little evidence that lowering systolic blood pressure, via antihypertensive drug classes, affected Alzheimer's disease risk but did not exclude the possibility of other mechanisms related to these drugs having an effect. This thesis has also examined the factors that have influenced prescribing trends of licensed dementia drugs, such as regulatory guidance and patient expiry. Understanding how existing drugs are used in this therapy area is key to maximising the benefit of repurposed drug candidates in the future.

References

1. Jobke B, McBride T, Nevin L, Peiperl L, Ross A, Stone C, et al. Setbacks in Alzheimer research demand new strategies, not surrender. PLOS Med. 2018 Feb 27;15(2):e1002518.
2. Gauthier S, Albert M, Fox N, Goedert M, Kivipelto M, Mestre-Ferrandiz J, et al. Why has therapy development for dementia failed in the last two decades? Alzheimers Dement. 2016 Jan 1;12(1):60–4.
3. Appleby BS, Cummings JL. Discovering new treatments for Alzheimer's disease by repurposing approved medications. Curr Top Med Chem. 2013;13(18):2306–27.
4. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004 Aug;3(8):673–83.
5. Corbett A, Pickett J, Burns A, Corcoran J, Dunnett SB, Edison P, et al. Drug repositioning for Alzheimer's disease. Nat Rev Drug Discov. 2012;11(11):833–846.
6. Kennelly SP, Lawlor BA, Kenny RA. Blood pressure and the risk for dementia—A double edged sword. Ageing Res Rev. 2009 Apr 1;8(2):61–70.
7. Qiu C, Winblad B, Fratiglioni L. The age-dependent relation of blood pressure to cognitive function and dementia. Lancet Neurol. 2005 Aug;4(8):487–99.
8. Abell JG, Kivimäki M, Dugravot A, Tabak AG, Fayosse A, Shipley M, et al. Association between systolic blood pressure and dementia in the Whitehall II cohort study: role of age, duration, and threshold used to define hypertension. Eur Heart J. 2018 Sep 1;39(33):3119–25.
9. Kehoe PG, Miners S, Love S. Angiotensins in Alzheimer's disease - friend or foe? Trends Neurosci. 2009 Dec;32(12):619–28.
10. Kehoe PG, Passmore PA. The renin-angiotensin system and antihypertensive drugs in Alzheimer's disease: current standing of the angiotensin hypothesis? J Alzheimers Dis JAD. 2012;30 Suppl 2:S251-268.
11. Wright AK, Kontopantelis E, Emsley R, Buchan I, Sattar N, Rutter MK, et al. Life Expectancy and Cause-Specific Mortality in Type 2 Diabetes: A Population-Based Cohort Study Quantifying Relationships in Ethnic Subgroups. Diabetes Care. 2017 Mar 1;40(3):338–45.

12. Larsson SC, Markus HS. Does Treating Vascular Risk Factors Prevent Dementia and Alzheimer's Disease? A Systematic Review and Meta-Analysis. J Alzheimers Dis. 2018 Jan 1;64(2):657–68.
13. Ohrui T, Matsui T, Yamaya M, Arai H, Ebihara S, Maruyama M, et al. Angiotensin-converting enzyme inhibitors and incidence of Alzheimer's disease in Japan. J Am Geriatr Soc. 2004 Apr;52(4):649–50.
14. Sink KM, Leng X, Williamson J, Kritchevsky SB, Yaffe K, Kuller L, et al. Angiotensin-converting enzyme inhibitors and cognitive decline in older adults with hypertension: results from the Cardiovascular Health Study. Arch Intern Med. 2009 Jul 13;169(13):1195–202.
15. Davies NM, Kehoe PG, Ben-Shlomo Y, Martin RM. Associations of anti-hypertensive treatments with Alzheimer's disease, vascular dementia, and other dementias. J Alzheimers Dis. 2011;26(4):699.
16. Li N-C, Lee A, Whitmer RA, Kivipelto M, Lawler E, Kazis LE, et al. Use of angiotensin receptor blockers and risk of dementia in a predominantly male population: prospective cohort analysis. BMJ. 2010 Jan 12;340(jan12 1):b5465–b5465.
17. Hsu C-Y, Huang C-C, Chan W-L, Huang P-H, Chiang C-H, Chen T-J, et al. Angiotensin-receptor blockers and risk of Alzheimer's disease in hypertension population--a nationwide cohort study. Circ J Off J Jpn Circ Soc. 2013;77(2):405–10.
18. Goh KL, Bhaskaran K, Minassian C, Evans SJW, Smeeth L, Douglas IJ. Angiotensin receptor blockers and risk of dementia: cohort study in UK Clinical Practice Research Datalink. Br J Clin Pharmacol. 2015 Feb 1;79(2):337–50.
19. Hwang D, Kim S, Choi H, Oh I-H, Kim BS, Choi HR, et al. Calcium-Channel Blockers and Dementia Risk in Older Adults - National Health Insurance Service - Senior Cohort (2002-2013). Circ J Off J Jpn Circ Soc. 2016 Oct 25;80(11):2336–42.
20. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol. 2016 Dec 1;45(6):1866–86.
21. Walker VM, Kehoe PG, Martin RM, Davies NM. Repurposing antihypertensive drugs for the prevention of Alzheimer's disease: a Mendelian Randomization study. Int J Epidemiol. 2019 Jul 4; Advance article.
22. Walker V, Davies N, Martin R, Kehoe P. Comparison of antihypertensive drug classes for dementia prevention. bioRxiv. 2019 Jan 12;517482.

23. Hernán MA, Robins JM. Instruments for Causal Inference: An Epidemiologist's Dream? Epidemiology. 2006 Jul;17(4):360–72.
24. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000 Aug 1;29(4):722–9.
25. Davies NM, Gunnell D, Thomas KH, Metcalfe C, Windmeijer F, Martin RM. Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. J Clin Epidemiol. 2013 Dec;66(12):1386–96.
26. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiol Drug Saf. 2010;19(6):537–54.
27. Brookhart MA, Wang P, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiol Camb Mass. 2006;17(3):268–75.
28. Davies NM, Davey Smith G, Windmeijer F, Martin RM. COX-2 Selective Nonsteroidal Anti-inflammatory Drugs and Risk of Gastrointestinal Tract Complications and Myocardial Infarction: An Instrumental Variable Analysis. Epidemiology. 2013 May;24(3):352–62.
29. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int J Epidemiol. 2014 Jun 1;43(3):922–9.
30. Freeman G, Cowling BJ, Schooling CM. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. Int J Epidemiol. 2013 Aug 1;42(4):1157–63.
31. Walker VM, Davies NM, Windmeijer F, Burgess S, Martin RM. Power calculator for instrumental variable analysis in pharmacoepidemiology. Int J Epidemiol. 2017 Oct 1;46(5):1627–32.
32. Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol. 2017 Dec 1;46(6):2078–89.
33. Mukadam N, Livingston G, Rantell K, Rickman S. Diagnostic rates and treatment of dementia before and after launch of a national dementia policy: an observational study using English national databases. BMJ Open. 2014 Jan 1;4(1):e004119.

34. Donegan K, Fox N, Black N, Livingston G, Banerjee S, Burns A. Trends in diagnosis and treatment for people with dementia in the UK from 2005 to 2015: a longitudinal retrospective cohort study. Lancet Public Health. 2017 Mar 1;2(3):e149–56.
35. Walker VM, Davies NM, Kehoe PG, Martin RM. What is the impact of regulatory guidance and expiry of drug patents on dementia drug prescriptions in England? A trend analysis in the Clinical Practice Research Datalink. Alzheimers Res Ther. 2018 May 29;10:51.
36. Walker VM, Davies NM, Jones T, Kehoe PG, Martin RM. Can commonly prescribed drugs be repurposed for the prevention or treatment of Alzheimer's and other neurodegenerative diseases? Protocol for an observational cohort study in the UK Clinical Practice Research Datalink. BMJ Open. 2016 Dec 1;6(12):e012044.
37. Prince M, Wimo A, Guerchet M, Ali G-C, Wu Y-T, Prina M. World Alzheimer Report 2015. London: Alzheimer's Disease International; 2015 Aug.
38. Alzheimer's Society. Dementia UK: Update [Internet]. 2014 [cited 2016 Feb 11]. Available from: https://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=2323
39. Deaths registered in England and Wales (series DR) - Office for National Statistics [Internet]. [cited 2018 Dec 12]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredinenglandandwalesseriesdr/2015#dementia-and-alzheimer-disease-was-the-leading-cause-of-death-in-2015
40. Deaths due to dementia | Dementia Statistics Hub [Internet]. [cited 2018 Dec 12]. Available from: https://www.dementiastatistics.org/statistics/deaths-due-to-dementia/
41. Calcoen D, Elias L, Yu X. What does it take to produce a breakthrough drug? Nat Rev Drug Discov. 2015 Mar;14(3):161–2.
42. Albanese E, Banerjee S, Dhanasiri S, Fernandez J-L, Ferri C, Knapp M, et al. Dementia UK [Internet]. 2007 [cited 2019 Feb 21]. Available from: https://www.alzheimers.org.uk/sites/default/files/2018-10/Dementia_UK_Full_Report_2007.pdf?fileID=2
43. Ballard C, Gauthier S, Corbett A, Brayne C, Aarsland D, Jones E. Alzheimer's Disease. The Lancet. 2011;377:1019–31.
44. Masters CL, Beyreuther K. Science, Medicine, and the Future: Alzheimer's Disease. BMJ. 1998;316(7129):446–8.

45. Whitmer RA, Sidney S, Selby J, Johnston SC, Yaffe K. Midlife cardiovascular risk factors and risk of dementia in late life. Neurology. 2005;64(2):277–281.
46. Kivipelto M, Helkala E-L, Laakso MP, Hänninen T, Hallikainen M, Alhainen K, et al. Midlife vascular risk factors and Alzheimer's disease in later life: longitudinal, population based study. BMJ. 2001;322(7300):1447–1451.
47. Strand BH, Langballe EM, Hjellvik V, Handal M, Næss Ø, Knudsen GP, et al. Midlife vascular risk factors and their association with dementia deaths: Results from a Norwegian prospective study followed up for 35years. J Neurol Sci. 2013 Jan;324(1–2):124–30.
48. British National Formulary [Internet]. MedicinesComplete. [cited 2018 Sep 11]. Available from: https://about.medicinescomplete.com/publication/british-national-formulary/
49. Dementia | Treatment summary | BNF content published by NICE [Internet]. [cited 2018 Dec 12]. Available from: https://bnf.nice.org.uk/treatment-summary/dementia.html
50. Alzheimer's Disease:Written question - 138987 - UK Parliament [Internet]. [cited 2018 Dec 12]. Available from: https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2018-04-27/138987
51. Ritchie CW, Molinuevo JL, Truyen L, Satlin A, Van der Geyten S, Lovestone S. Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project. Lancet Psychiatry. 2016 Feb;3(2):179–86.
52. NHS Choices. Medicines information - Licensing [Internet]. 2015 [cited 2015 Dec 16]. Available from: http://www.nhs.uk/Conditions/Medicinesinfo/Pages/Safetyissues.aspx
53. Waller PC, Evans SJW. A model for the future conduct of pharmacovigilance. Pharmacoepidemiol Drug Saf. 2003 Jan;12(1):17–29.
54. Pinnow E, Sharma P, Parekh A, Gevorkian N, Uhl K. Increasing participation of women in early phase clinical trials approved by the FDA. Womens Health Issues Off Publ Jacobs Inst Womens Health. 2009 Apr;19(2):89–93.

55. NatCen Social Research, University College London. Health Survey for England 2014 - Trend Table Commentary [Internet]. 2015 [cited 2016 Feb 23]. Available from: http://www.hscic.gov.uk/catalogue/PUB19297/HSE2014-Trend-commentary.pdf
56. Carretero Oscar A., Oparil Suzanne. Essential Hypertension. Circulation. 2000 Jan 25;101(3):329–35.
57. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. Lancet Neurol. 2011 Sep 1;10(9):819–28.
58. National Institute for Health and Care Excellence. Hypertension in adults: diagnosis and management. [Internet]. 2011 [cited 2017 Jun 2]. Available from: http://www.nice.org.uk/guidance/cg127/resources/hypertension-in-adults-diagnosis-and-management-35109454941637
59. Law MR, Wald NJ, Morris JK, Jordan RE. Value of low dose combination treatment with blood pressure lowering drugs: analysis of 354 randomised trials. BMJ. 2003 Jun 26;326(7404):1427.
60. Rollin-Sillaire A, Breuilh L, Salleron J, Bombois S, Cassagnaud P, Deramecourt V, et al. Reasons that prevent the inclusion of Alzheimer's disease patients in clinical trials. Br J Clin Pharmacol. 2013 Apr;75(4):1089–97.
61. Schneider LS. The Potential and Limits for Clinical Trials for Early Alzheimer's Disease and Some Recommendations. J Nutr Health Aging. 2010 Apr;14(4):295–8.
62. Barthold D, Joyce G, Wharton W, Kehoe P, Zissimopoulos J. The association of multiple anti-hypertensive medication classes with Alzheimer's disease incidence across sex, race, and ethnicity. PLOS ONE. 2018 Nov 1;13(11):e0206705.
63. The SPRINT MIND Investigators for the SPRINT Research Group, Williamson JD, Pajewski NM, Auchus AP, Bryan RN, Chelune G et al. Effect of intensive vs standard blood pressure control on probable dementia: a randomized clinical trial. JAMA. 2019;321:553.
64. Robins JM. Data, design, and background knowledge in etiologic inference. Epidemiology. 2001;12(3):313–320.
65. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. Int J Epidemiol. 2018 Feb 1;47(1):226–35.

66. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to conditioning on a collider. Int J Epidemiol. 2010 Apr 1;39(2):417–20.
67. Suissa S. Immortal time bias in observational studies of drug effects. Pharmacoepidemiol Drug Saf. 2007 Mar;16(3):241–9.
68. Suissa S. Immortal Time Bias in Pharmacoepidemiology. Am J Epidemiol. 2008 Jan 7;167(4):492–9.
69. Levesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. Br Med J. 2010 Mar 12;340(mar12 1):b5087–b5087.
70. Mamdani M, Juurlink DN, Kopp A, Naglie G, Austin PC, Laupacis A. Gastrointestinal bleeding after the introduction of COX 2 inhibitors: ecological study. BMJ. 2004 Jun 10;328(7453):1415–6.
71. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. Nat Rev Rheumatol. 2015 Jul;11(7):437–41.
72. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne J a. C. Methods for dealing with time-dependent confounding. Stat Med. 2013 Apr 30;32(9):1584–618.
73. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016 Apr 15;183(8):758–64.
74. Hernán M, Robins J. Causal Inference. Boca Raton: Chapman & Hall/CRC; forthcoming.
75. Swanson SA, Miller M, Robins JM, Hernán MA. Definition and Evaluation of the Monotonicity Condition for Preference-Based Instruments. Epidemiol Camb Mass. 2015 May;26(3):414–20.
76. Schneeweiss S, Setoguchi S, Brookhart A, Dormuth C, Wang PS. Risk of death associated with the use of conventional versus atypical antipsychotic drugs among elderly patients. CMAJ Can Med Assoc J J Assoc Medicale Can. 2007 Feb 27;176(5):627–32.
77. Wang PS, Schneeweiss S, Avorn J, Fischer MA, Mogun H, Solomon DH, et al. Risk of death in elderly users of conventional vs. atypical antipsychotic medications. N Engl J Med. 2005 Dec 1;353(22):2335–41.
78. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. N Engl J Med. 2008 Feb 21;358(8):771–83.

79. Taylor GM, Taylor AE, Thomas KH, Jones T, Martin RM, Munafò MR, et al. The effectiveness of varenicline versus nicotine replacement therapy on long-term smoking cessation in primary care: a prospective cohort study of electronic medical records. Int J Epidemiol. 2017 Dec 1;46(6):1948–57.
80. StataCorp. Stata Statistical Software. College Station, TX: StataCorp LP; 2015.
81. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003 Feb 1;32(1):1–22.
82. Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. Int J Epidemiol. 2004 Feb 1;33(1):30–42.
83. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014 Sep 15;23(R1):R89–98.
84. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008 Apr 15;27(8):1133–63.
85. Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. Int J Epidemiol. 2016 Jun;45(3):908–15.
86. Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. Int J Epidemiol. 2016 Dec 1;45(6):1717–26.
87. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med. 2017 Mar 29;9(383).
88. Tuffs A. Bayer faces potential fine over cholesterol lowering drug. BMJ. 2001 Aug 25;323(7310):415.
89. Furberg CD, Pitt B. Withdrawal of cerivastatin from the world market. Curr Control Trials Cardiovasc Med. 2001;2(5):205–7.
90. Staffa JA, Chang J, Green L. Cerivastatin and Reports of Fatal Rhabdomyolysis. N Engl J Med. 2002 Feb 14;346(7):539–40.
91. Collins R, Reith C, Emberson J, Armitage J, Baigent C, Blackwell L, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. The Lancet. 2016 Nov 19;388(10059):2532–61.

92. Graham DJ, Staffa JA, Shatin D, Andrade SE, Schech SD, La Grenade L, et al. Incidence of Hospitalized Rhabdomyolysis in Patients Treated with Lipid-Lowering Drugs. JAMA. 2004 Dec 1;292(21):2585–90.
93. Bonetti PO, Lerman LO, Napoli C, Lerman A. Statin effects beyond lipid lowering—are they clinically relevant? Eur Heart J. 2003 Feb 1;24(3):225–48.
94. Laufs U, Wassmann S, Hilgers S, Ribaudo N, Böhm M, Nickenig G. Rapid effects on vascular function after initiation and withdrawal of atorvastatin in healthy, normocholesterolemic men. Am J Cardiol. 2001 Dec 1;88(11):1306–7.
95. Tsunekawa T, Hayashi T, Kano H, Sumi D, Matsui-Hirai H, Thakur NK, et al. Cerivastatin, a hydroxymethylglutaryl coenzyme a reductase inhibitor, improves endothelial function in elderly diabetic patients within 3 days. Circulation. 2001 Jul 24;104(4):376–9.
96. Davignon J, Laaksonen R. Low-density lipoprotein-independent effects of statins. Curr Opin Lipidol. 1999 Dec;10(6):543–59.
97. Lefer AM, Scalia R, Lefer DJ. Vascular effects of HMG CoA-reductase inhibitors (statins) unrelated to cholesterol lowering: new concepts for cardiovascular disease. Cardiovasc Res. 2001 Feb 1;49(2):281–7.
98. Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes MV, Engmann JEL, Shah T, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. The Lancet. 2015 Jan 24;385(9965):351–61.
99. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet. 2008 Feb;40(2):189–97.
100. Sattar N, Preiss D, Murray HM, Welsh P, Buckley BM, de Craen AJ, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. The Lancet. 2010 Mar 5;375(9716):735–42.
101. Lotta LA, Sharp SJ, Burgess S, Perry JRB, Stewart ID, Willems SM, et al. Association Between Low-Density Lipoprotein Cholesterol-Lowering Genetic Variants and Risk of Type 2 Diabetes: A Meta-analysis. JAMA. 2016 Oct 4;316(13):1383–91.
102. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. N Engl J Med. 2016 Dec 1;375(22):2144–53.

103. Yin P, Anttila V, Siewert KM, Palotie A, Davey Smith G, Voight BF. Serum calcium and risk of migraine: a Mendelian randomization study. Hum Mol Genet. 2017 Feb 15;26(4):820–8.
104. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018 May 30;7:e34408.
105. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, EPIC-InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol. 2015 Jul;30(7):543–52.
106. Taylor AE, Davies NM, Ware JJ, VanderWeele T, Davey Smith G, Munafò MR. Mendelian randomization in health research: Using appropriate genetic variants and avoiding biased estimates. Econ Hum Biol. 2014 Mar 1;13:99–106.
107. Millard LAC, Davies NM, Timpson NJ, Tilling K, Flach PA, Davey Smith G. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. Sci Rep. 2015 Nov 16;5:16645.
108. The Interleukin 1 Genetics Consortium. Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis. Lancet Diabetes Endocrinol. 2015 Apr;3(4):243–53.
109. Matteini AM, Li J, Lange EM, Tanaka T, Lange LA, Tracy RP, et al. Novel gene variants predict serum levels of the cytokines IL-18 and IL-1ra in older adults. Cytokine. 2014 Jan;65(1):10–6.
110. Novartis Phase III study shows ACZ885 (canakinumab) reduces cardiovascular risk in people who survived a heart attack [Internet]. Novartis. [cited 2017 Jul 27]. Available from: https://www.novartis.com/news/media-releases/novartis-phase-iii-study-shows-acz885-canakinumab-reduces-cardiovascular-risk
111. Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. Annu Rev Genomics Hum Genet. 2016;17(1):353–73.
112. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. Nat Rev Genet. 2016 Mar;17(3):129–45.
113. Walker AM. Confounding by Indication. Epidemiology. 1996 Jul;7(4):335–6.

114. Brookhart M, Stürmer T, Glynn R, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. Med Care. 2010 Jun;48(6 0):S114–20.
115. Davey Smith G. Use of genetic markers and gene-diet interactions for interrogating population-level causal influences of diet on health. Genes Nutr. 2011 Feb;6(1):27–43.
116. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered Environments and Randomized Genes: A Fundamental Distinction between Conventional and Genetic Epidemiology. PLOS Med. 2007 Dec 11;4(12):e352.
117. Groenwold RHH, Hak E, Hoes AW. Quantitative assessment of unobserved confounding is mandatory in nonrandomized intervention studies. J Clin Epidemiol. 2009 Jan;62(1):22–8.
118. Stampfer MJ, Colditz GA. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. Prev Med. 1991 Jan;20(1):47–63.
119. Petitti D. Commentary: Hormone replacement therapy and coronary heart disease: four lessons. Int J Epidemiol. 2004 Jun 1;33(3):461–3.
120. Lawlor DA, Davey Smith G, Ebrahim S. Socioeconomic Position and Hormone Replacement Therapy Use: Explaining the Discrepancy in Evidence From Observational and Randomized Controlled Trials. Am J Public Health. 2004 Dec 1;94(12):2149–54.
121. Yeung SLA, Cheng KK, Zhao J, Zhang W, Jiang C, Lam TH, et al. Genetically predicted 17beta-estradiol and cardiovascular risk factors in women: a Mendelian randomization analysis using young women in Hong Kong and older women in the Guangzhou Biobank Cohort Study. Ann Epidemiol. 2016 Mar 1;26(3):171–5.
122. Yarmolinsky J, Bonilla C, Haycock PC, Langdon RJQ, Lotta LA, Langenberg C, et al. Circulating Selenium and Prostate Cancer Risk: A Mendelian Randomization Analysis. J Natl Cancer Inst. 2018 01;110(9):1035–8.
123. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. N Engl J Med. 2017 May 4;376(18):1713–22.
124. Silverman MG, Ference BA, Im K, Wiviott SD, Giugliano RP, Grundy SM, et al. Association Between Lowering LDL-C and Cardiovascular Risk Reduction Among

Different Therapeutic Interventions: A Systematic Review and Meta-analysis. JAMA. 2016 Sep 27;316(12):1289–97.
125. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. The Lancet. 2012 Aug 11;380(9841):572–80.
126. Schmidt AF, Swerdlow DI, Holmes MV, Patel RS, Fairhurst-Hunter Z, Lyall DM, et al. PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2017 Feb;5(2):97–105.
127. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel Data-Mining Methodologies for Adverse Drug Event Discovery and Analysis. Clin Pharmacol Ther. 2012 Jun;91(6):1010–21.
128. Last JM, editor. Berkson's bias. In: A Dictionary of Public Health [Internet]. Oxford University Press; 2007 [cited 2018 Feb 27]. Available from: http://www.oxfordreference.com/view/10.1093/acref/9780195160901.001.0001/acref-9780195160901-e-392
129. Westreich D. Berkson's bias, selection bias, and missing data. Epidemiol Camb Mass. 2012 Jan;23(1):159–64.
130. Graves EJ, Gillum BS. Detailed diagnoses and procedures, National Hospital Discharge Survey, 1995. Vital Health Stat 13. 1997 Nov;(130):1–146.
131. Chavez LO, Leon M, Einav S, Varon J. Beyond muscle destruction: a systematic review of rhabdomyolysis for clinical practice. Crit Care. 2016 June;20(1):135.
132. Torres PA, Helmstetter JA, Kaye AM, Kaye AD. Rhabdomyolysis: pathogenesis, diagnosis, and treatment. Ochsner J. 2015 Spring;15(1):58–69.
133. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016 May;40(4):304–14.
134. Omar MA, Wilson JP, Cox TS. Rhabdomyolysis and HMG-CoA reductase inhibitors. Ann Pharmacother. 2001 Sep;35(9):1096–107.
135. Khan FY. Rhabdomyolysis: a review of the literature. Neth J Med. 2009 Oct;67(9):272–83.
136. Joy TR. Narrative Review: Statin-Related Myopathy. Ann Intern Med. 2009 Jun 16;150(12):858.

137. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015 Apr;44(2):512–25.
138. Holmes MV, Ala-Korpela M, Davey Smith G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat Rev Cardiol. 2017 Oct;14(10):577–90.
139. Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors. J Clin Epidemiol. 2016 Jan;69:208–16.
140. Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. PLoS Genet. 2017 Oct;13(10):e1006944.
141. Dudbridge F, Allen RJ, Sheehan NA, Schmidt AF, Lee JC, Jenkins RG, et al. Adjustment for index event bias in genome-wide association studies of subsequent events. Nat Commun. 2019 Apr 5;10(1):1–10.
142. Bound J, Jaeger DA, Baker RM. Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. J Am Stat Assoc. 1995;90(430):443–50.
143. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. J Clin Epidemiol. 2009 Dec;62(12):1226–32.
144. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application—in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. J Clin Epidemiol. 2009 Dec;62(12):1233–41.
145. Davies NM, von Hinke Kessler Scholder S, Farbmacher H, Burgess S, Windmeijer F, Davey Smith G. The many weak instruments problem and Mendelian randomization. Stat Med. 2015 Feb 10;34(3):454–68.
146. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? Hum Genet. 2008 Feb;123(1):15–33.

147. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015 Jun;44(3):827–36.
148. CPRD Governance [Internet]. [cited 2015 Nov 17]. Available from: http://www.cprd.com/governance/
149. Walker V, Davies N, Kehoe P, Martin R. CPRD codes: ICD-10 equivalent code lists for dementia subtypes. 2017; Available from: https://doi.org/10.5523/bris.2h4rmk9v7pw2k23h7vgf9tx1ea
150. Walker V, Davies N, Kehoe P, Martin R. CPRD codes: neurodegenerative diseases and commonly prescribed drugs. 2017; Available from: https://doi.org/10.5523/bris.1plm8il42rmlo2a2fqwslwckm2
151. Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. J Public Health. 2014 Dec 1;36(4):684–92.
152. Bhaskaran K, Forbes HJ, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). BMJ Open. 2013 Sep 13;3(9):e003389–e003389.
153. Campbell J, Dedman D, Eaton S, Gallagher A, Williams T. Is the CPRD GOLD population comparable to the UK population. Pharmacoepidemiol Drug Saf. 2013;22(Suppl 1):280–1.
154. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract. 2010 Mar 1;60(572):128–36.
155. McGuinness LA, Warren-Gash C, Moorhouse LR, Thomas SL. The validity of dementia diagnoses in routinely collected electronic health records in the United Kingdom: A systematic review. Pharmacoepidemiol Drug Saf. 2019 Feb;28(2):244–55.
156. García Rodríguez LA, Pérez Gutthann S. Use of the UK general practice research database for pharmacoepidemiology. Br J Clin Pharmacol. 1998;45(5):419–425.
157. University of Manchester. ClinicalCodes Repository [Internet]. [cited 2016 Feb 5]. Available from: https://clinicalcodes.rss.mhs.man.ac.uk/
158. Thomas KH, Davies N, Metcalfe C, Windmeijer F, Martin RM, Gunnell D. Validation of suicide and self-harm records in the Clinical Practice Research Datalink. Br J Clin Pharmacol. 2013 Jul;76(1):145–57.

159. Health & Social Care Information Centre. Read Codes [Internet]. [cited 2016 Dec 16]. Available from: http://webarchive.nationalarchives.gov.uk/20160921135209/http://systems.digital.nhs.uk/data/uktc/readcodes
160. BNF Legacy. Section 4.11 : Drugs for Dementia [Internet]. 2015 [cited 2016 Jan 11]. Available from: https://www.medicinescomplete.com/mc/bnflegacy/current/PHP3236-drugs-for-dementia.htm
161. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975 Nov;12(3):189–98.
162. National Institute for Health and Care Excellence. Press Release: NICE consults on revised first draft guidance on the use of drugs to treat Alzheimer's disease (including a review of existing guidance no. 19) [Internet]. [cited 2016 Nov 24]. Available from: https://www.nice.org.uk/guidance/ta111/resources/2006001-nice-consults-on-revised-first-draft-guidance-on-the-use-of-drugs-to-treat-alzheimers-disease2
163. National Institute for Health and Care Excellence. Alzheimer's disease (mild to moderate) - donepezil, galantamine, rivastigmine and memantine (part review): final scope [Internet]. 2009 [cited 2017 Feb 6]. Available from: https://www.nice.org.uk/guidance/TA217/documents/alzheimers-disease-mild-to-moderate-donepezil-galantamine-rivastigmine-and-memantine-part-review-final-scope2
164. National Institute for Health and Care Excellence. Dementia: supporting people with dementia and their carers in health and social care [Internet]. 2006 [cited 2017 Jun 2]. Available from: http://www.nice.org.uk/guidance/cg42/resources/dementia-supporting-people-with-dementia-and-their-carers-in-health-and-social-care-975443665093
165. Dyer C. NICE's decision on dementia drugs was "irrational," High Court is told. BMJ. 2007 Jun 30;334(7608):1337.
166. Kmietowicz Z. NICE hears appeals over dementia drugs. BMJ. 2006 Jul 22;333(7560):165.
167. Iliffe S. The National Institute for Health and Clinical Excellence (NICE) and Drug Treatment for Alzheimer's Disease. CNS Drugs. 2007 Mar 1;21(3):177–84.

168. National Institute for Health and Care Excellence. Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease. NICE technology appraisal guidance 217 [Internet]. 2011 [cited 2017 May 3]. Available from: https://www.nice.org.uk/guidance/ta217
169. NHS Digital 1 Trevelyan Square. Quality and Outcomes Framework [Internet]. 2015 [cited 2017 Jan 18]. Available from: http://content.digital.nhs.uk/qof
170. Information Centre, Government Statistical Service. National Quality and Outcomes Framework Statistics for England 2006/07 [Internet]. 2007 [cited 2017 Jan 11]. Available from: http://content.digital.nhs.uk/catalogue/PUB05997/qof-eng-06-07-bull-rep.pdf
171. NHS Digital. Summary of QOF indicators [Internet]. [cited 2017 Jan 18]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213226/Summary-of-QOF-indicators.pdf
172. Department of Health. Living well with dementia: A National Dementia Strategy [Internet]. 2009 Feb [cited 2016 Sep 14]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/168220/dh_094051.pdf
173. Department of Health. Prime Minister's challenge on dementia - delivering major improvements in dementia care and research by 2015 [Internet]. 2012 [cited 2016 Dec 16]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/215101/dh_133176.pdf
174. Department of Health. Prime Minister's challenge on dementia 2020 [Internet]. 2015 [cited 2016 Dec 16]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/414344/pm-dementia2020.pdf
175. Abdi Z, Burns A. Championing of dementia in England. Alzheimers Res Ther. 2012;4(5):36.
176. Wortmann M. Importance of national plans for Alzheimer's disease and dementia. Alzheimers Res Ther. 2013;5(5):40.
177. Alderwick H, Robertson R, Appleby J, Dunn P, Maguire D. Better value in the NHS: The role of changes in clinical practice [Internet]. 2015 [cited 2016 Dec 13].

Available from:
https://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/better-value-nhs-Kings-Fund-July%202015.pdf
178. Boarer C, Solomons K. Acetylcholinesterase inhibitors: maximising benefits to the Surrey healthcare economy from the loss of exclusivity of donepezil, galantamine and rivastigmine in 2012-13 [Internet]. 2012 [cited 2016 Dec 16]. Available from: http://pad.res360.net/Content/Documents/Loss%20of%20patent%20exclusivity%20of%20acetylcholinesterase%20inhibitors.pdf
179. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med. 2000 Feb 15;19(3):335–51.
180. Statistical Methodology and Applications Branch, Surveillance Research Program, National Cancer Institute. Dependent Variable — Joinpoint Help System 4.5.0.1 [Internet]. [cited 2017 Nov 20]. Available from: https://surveillance.cancer.gov/help/joinpoint/setting-parameters/input-file-tab/dependent-variable
181. Statistical Methodology and Applications Branch, Surveillance Research Program, National Cancer Institute. Joinpoint Regression Program. 2016.
182. Medicine & Healthcare products Regulatory Agency, National Institute for Health Research, Clinical Practice Research Datalink. Hospital Episode Statistics (HES) Outpatient Care and GOLD Documentation (Set 12). 2016.
183. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Contin Educ Anaesth Crit Care Pain. 2008 Dec 1;8(6):221–3.
184. Google. Interest over time on Google Trends for Alzheimer's disease - England, 1/1/08 - 12/1/15 [Internet]. [cited 2018 Apr 4]. Available from: https://g.co/trends/TTrq9
185. Buckley JS, Salpeter SR. A Risk-Benefit Assessment of Dementia Medications: Systematic Review of the Evidence. Drugs Aging. 2015 Jun 1;32(6):453–67.
186. Rodda J, Carter J. Cholinesterase inhibitors and memantine for symptomatic treatment of dementia. BMJ. 2012 May 1;344:e2986.
187. Glossary [Internet]. NICE. [cited 2018 Apr 4]. Available from: https://www.nice.org.uk/glossary?letter=f

188. Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. Curr Epidemiol Rep. 2015 Dec;2(4):221–8.
189. BNF Legacy. BNF July 2017: BNF Legacy [Internet]. [cited 2018 Feb 16]. Available from: https://www.medicinescomplete.com/mc/bnflegacy/64/
190. Hernán M, Alonso A, Logan R, Grodstein F, Michels K, Willett W, et al. Observational Studies Analyzed Like Randomized Experiments. Epidemiology. 2008 Nov 1;19(6):766–79.
191. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. Int J Biostat. 2007;3(1):14.
192. Baum CF, Schaffer ME, Stillman S. ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. [Internet]. 2010 [cited 2017 Nov 9]. Available from: http://ideas.repec.org/c/boc/bocode/s425401.html
193. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. J Am Stat Assoc. 1996 Jun;91(434):444–55.
194. Davies NM, Davey Smith G, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: a systematic review. Epidemiol Camb Mass. 2013 May;24(3):363–9.
195. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; Available from: https://www.R-project.org
196. Jackson JW, Swanson SA. Toward a Clearer Portrayal of Confounding Bias in Instrumental Variable Applications: Epidemiology. 2015 Jul;26(4):498–504.
197. Davies NM, Thomas KH, Taylor AE, Taylor GM, Martin RM, Munafò MR, et al. How to compare instrumental variable and conventional regression analyses using negative controls and bias plots. Int J Epidemiol. 2017 Dec 1;46(6):2067–77.
198. The National Institute for Health and Care Excellence. Beta-adrenoceptor blocking drugs [Internet]. [cited 2018 Mar 19]. Available from: https://bnf.nice.org.uk/treatment-summary/beta-adrenoceptor-blocking-drugs.html
199. CPRD @ Cambridge - Code Lists [Internet]. Primary Care Unit. [cited 2018 Mar 19]. Available from: http://www.phpc.cam.ac.uk/pcu/cprd_cam/codelists/

200. Kehoe PG, Blair PS, Howden B, Thomas DL, Malone IB, Horwood J, et al. The Rationale and Design of the Reducing Pathology in Alzheimer's Disease through Angiotensin TaRgeting (RADAR) Trial. J Alzheimers Dis JAD. 2018;61(2):803–14.
201. Mukund Lad. The English Indices of Deprivation 2010 [Internet]. 2011 [cited 2019 Jan 22]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/6871/1871208.pdf
202. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373–83.
203. Khan NF, Perera R, Harper S, Rose PW. Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. BMC Fam Pract. 2010 Jan 5;11:1.
204. Østergaard SD, Mukherjee S, Sharp SJ, Proitsi P, Lotta LA, Day F, et al. Associations between Potentially Modifiable Risk Factors and Alzheimer Disease: A Mendelian Randomization Study. PLOS Med. 2015 Jun 16;12(6):e1001841.
205. Larsson SC, Traylor M, Malik R, Dichgans M, Burgess S, Markus HS. Modifiable pathways in Alzheimer's disease: Mendelian randomisation analysis. BMJ. 2017 Dec 7;359:j5375.
206. Gill D, Georgakis MK, Koskeridis F, Jiang L, Feng Q, Wei W-Q, et al. Use of Genetic Variants Related to Antihypertensive Drugs to Inform on Efficacy and Side Effects. Circulation. 2019 Jun 25;140(4):270–9.
207. Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK Biobank [Internet]. Neale lab. [cited 2018 Aug 1]. Available from: http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank
208. Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013 Dec;45(12):1452–8.
209. Allen NE, Sudlow C, Peakman T, Collins R, UK Biobank. UK biobank data: come and get it. Sci Transl Med. 2014 Feb 19;6(224):224ed4.
210. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank

Participants With Those of the General Population. Am J Epidemiol. 2017 Nov 1;186(9):1026–34.
211. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018 Jan 4;46(D1):D1074–82.
212. GTEx Consortium, Analysts L, Laboratory DA& CC (LDACC), Management N program, Collection B, Pathology, et al. Genetic effects on gene expression across human tissues. Nature. 2017 Oct 12;550(7675):204–13.
213. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015 Oct;526(7571):68–74.
214. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013 Nov;37(7):658–65.
215. Burgess S, Bowden J. Integrating summarized data from multiple genetic variants in Mendelian randomization: bias and coverage properties of inverse-variance weighted methods. arXiv. 2015 Nov 27;04486.
216. Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. Int J Epidemiol. 2016 Oct;45(5):1600–16.
217. Kehoe PG. The coming of age of the angiotensin hypothesis in Alzheimer's disease: Progress toward disease prevention and treatment? J Alzheimers Dis. 2018 Mar;62(3):1443–66.
218. Yasar S, Xia J, Yao W, Furberg CD, Xue Q-L, Mercado CI, et al. Antihypertensive drugs decrease risk of Alzheimer disease: Ginkgo Evaluation of Memory Study. Neurology. 2013 Sep 3;81(10):896–903.
219. Davey Smith G, Ebrahim S. National Research Council (US) Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Survey. Mendelian Randomization: Genetic Variants as Instruments for Strengthening Causal Inference in Observational Studies. National Academies Press (US); 2008.
220. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015 Aug;47(8):856–60.

221. Nelson MR, Johnson T, Warren L, Hughes AR, Chissoe SL, Xu C-F, et al. The genetics of drug efficacy: opportunities and challenges. Nat Rev Genet. 2016 Apr;17(4):197–206.
222. Hurle MR, Nelson MR, Agarwal P, Cardon LR. Trial watch: Impact of genetically supported target selection on R&D productivity. Nat Rev Drug Discov. 2016 Sep;15(9):596–7.
223. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct;562(7726):203–9.
224. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol. 2016 Feb;70:214–23.
225. Bradford Hill A. The Environment and Disease: Association or Causation? Proc R Soc Med. 1965 May;58(5):295–300.
226. Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors. Journal of Clinical Epidemiology. 2016 Jan;69:208–16.

Appendix A. Contributions to scientific literature

This appendix contains the following papers:

Walker VM, Davies NM, Jones T, Kehoe PG, Martin RM. Can commonly prescribed drugs be repurposed for the prevention or treatment of Alzheimer's and other neurodegenerative diseases? Protocol for an observational cohort study in the UK Clinical Practice Research Datalink. BMJ Open. 2016 Dec 1;6(12):e012044.

Walker VM, Davies NM, Windmeijer F, Burgess S, Martin RM. Power calculator for instrumental variable analysis in pharmacoepidemiology. Int J Epidemiol. 2017 Oct 1;46(5):1627–32.

Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol. 2017 Dec 1;46(6):2078–89.

Walker VM, Davies NM, Kehoe PG, Martin RM. What is the impact of regulatory guidance and expiry of drug patents on dementia drug prescriptions in England? A trend analysis in the Clinical Practice Research Datalink. Alzheimer's Research & Therapy. 2018 May 29;10:51.

Walker VM, Davies NM, Martin RM, Kehoe PG. Comparison of antihypertensive drug classes for dementia prevention. bioRxiv. 2019 Jan 12;517482.

Walker VM, Kehoe PG, Martin RM, Davies NM. Repurposing antihypertensive drugs for the prevention of Alzheimer's disease: a Mendelian Randomization study. Int J Epidemiol. 2019 Jul 4; Advance article.

Appendix B. CPRD search terms

Table A.1: Medical code search terms.

Medical Event	Search Term
Alzheimer's disease	*alzheimer*
Amyotrophic lateral sclerosis	*amyotrophic lateral sclerosis*
Amyotrophic lateral sclerosis	*motor neur*
Dementia	*dementia*
Diabetes	*glucose*
Diabetes	*haemoglobin*A1c*
Diabetes	*HbA1c*
Diabetes	*impaired*glycemia*
Diabetes	*prediabetes*
Diabetes	*type 2 diabetes*
Diabetes	*type II diabetes*
Diabetes	*diabet*
Hypercholesterolaemia	*cholesterol*
Hypercholesterolaemia	*LDL*
Hypercholesterolaemia	*HDL*
Hypertension	*blood pressure*
Hypertension	*hypertension*
Parkinson's Disease	*parkinson*

Table A.2: Product code search terms.

Drug Category	Drug Subclass	Drug Substance	Search Term
Hypertension	Beta-adrenoceptor blockers	Propranolol hydrochloride	*propranolol*
Hypertension	Beta-adrenoceptor blockers	Acebutolol	*acebutolol*
Hypertension	Beta-adrenoceptor blockers	Atenolol	*atenolol*
Hypertension	Beta-adrenoceptor blockers	Bisoprolol fumarate	*bisoprolol*
Hypertension	Beta-adrenoceptor blockers	Carvedilol	*carvedilol*
Hypertension	Beta-adrenoceptor blockers	Celiprolol hydrochloride	*celiprolol*
Hypertension	Beta-adrenoceptor blockers	Co-tenidone	*tenidone*
Hypertension	Beta-adrenoceptor blockers	Labetalol hydrochloride	*labetalol*
Hypertension	Beta-adrenoceptor blockers	Metoprolol tartrate	*metoprolol*
Hypertension	Beta-adrenoceptor blockers	Nadolol	*nadolol*
Hypertension	Beta-adrenoceptor blockers	Nebivolol	*nebivolol*
Hypertension	Beta-adrenoceptor blockers	Oxprenolol hydrochloride	*oxprenolol*
Hypertension	Beta-adrenoceptor blockers	Pindolol	*pindolol*
Hypertension	Beta-adrenoceptor blockers	Sotalol hydrochloride	*sotalol*
Hypertension	Beta-adrenoceptor blockers	Timolol maleate	*timolol*
Hypertension	Angiotensin-converting enzyme inhibitors	Captopril	*captopril*
Hypertension	Angiotensin-converting enzyme inhibitors	Cilazapril	*cilazapril*
Hypertension	Angiotensin-converting enzyme inhibitors	Trandolapril	*trandolapril*
Hypertension	Angiotensin-converting enzyme inhibitors	Ramipril	*ramipril*
Hypertension	Angiotensin-converting enzyme inhibitors	Lisinopril	*lisinopril*
Hypertension	Angiotensin-converting enzyme inhibitors	Enalapril maleate	*enalapril*
Hypertension	Angiotensin-converting enzyme inhibitors	Fosinopril sodium	*fosinopril*
Hypertension	Angiotensin-converting enzyme inhibitors	Imidapril hydrochloride	*imidapril*
Hypertension	Angiotensin-converting enzyme inhibitors	Moexipril hydrochloride	*moexipril*
Hypertension	Angiotensin-converting enzyme inhibitors	Perindopril arginine	*perindopril*

Hypertension	Angiotensin-converting enzyme inhibitors	Perindopril erbumine	*perindopril*
Hypertension	Angiotensin-converting enzyme inhibitors	Quinapril	*quinapril*
Hypertension	Thiazides and related diuretics	Bendroflumethiazide	*bendroflumethiazide*
Hypertension	Thiazides and related diuretics	Chlortalidone	*chlortalidone*
Hypertension	Thiazides and related diuretics	Cyclopenthiazide	*cyclopenthiazide*
Hypertension	Thiazides and related diuretics	Indapamide	*indapamide*
Hypertension	Thiazides and related diuretics	Metolazone	*metolazone*
Hypertension	Thiazides and related diuretics	Xipamide	*xipamide*
Hypertension	Calcium channel blockers	Amlodipine	*amlodipine*
Hypertension	Calcium channel blockers	Diltiazem hydrochloride	*diltiazem*
Hypertension	Calcium channel blockers	Felodipine	*felodipine*
Hypertension	Calcium channel blockers	Isradipine	*isradipine*
Hypertension	Calcium channel blockers	Lacidipne	*lacidipine*
Hypertension	Calcium channel blockers	Lercanidipine hydrochloride	*lercanidipine*
Hypertension	Calcium channel blockers	Nicardipine hydrochloride	*nicardipine*
Hypertension	Calcium channel blockers	Nifedipine	*nifedipine*
Hypertension	Calcium channel blockers	Nimodipine	*nimodipine*
Hypertension	Calcium channel blockers	Verapamil hydrochloride	*verapamil*
Hypertension	Loop diuretics	Bumetanide	*bumetanide*
Hypertension	Loop diuretics	Furosemide	*furosemide*
Hypertension	Loop diuretics	Torasemide	*torasemide*
Hypertension	Alpha-adrenoceptor blockers	Doxazosin	*doxazosin*
Hypertension	Alpha-adrenoceptor blockers	Indoramin	*indoramin*
Hypertension	Alpha-adrenoceptor blockers	Phenoxybenzamine hydrochloride	*phenoxybenzamine*
Hypertension	Alpha-adrenoceptor blockers	Phentolamine mesilate	*phentolamine*
Hypertension	Alpha-adrenoceptor blockers	Prazosin	*prazosin*
Hypertension	Alpha-adrenoceptor blockers	Terazosin	*terazosin*

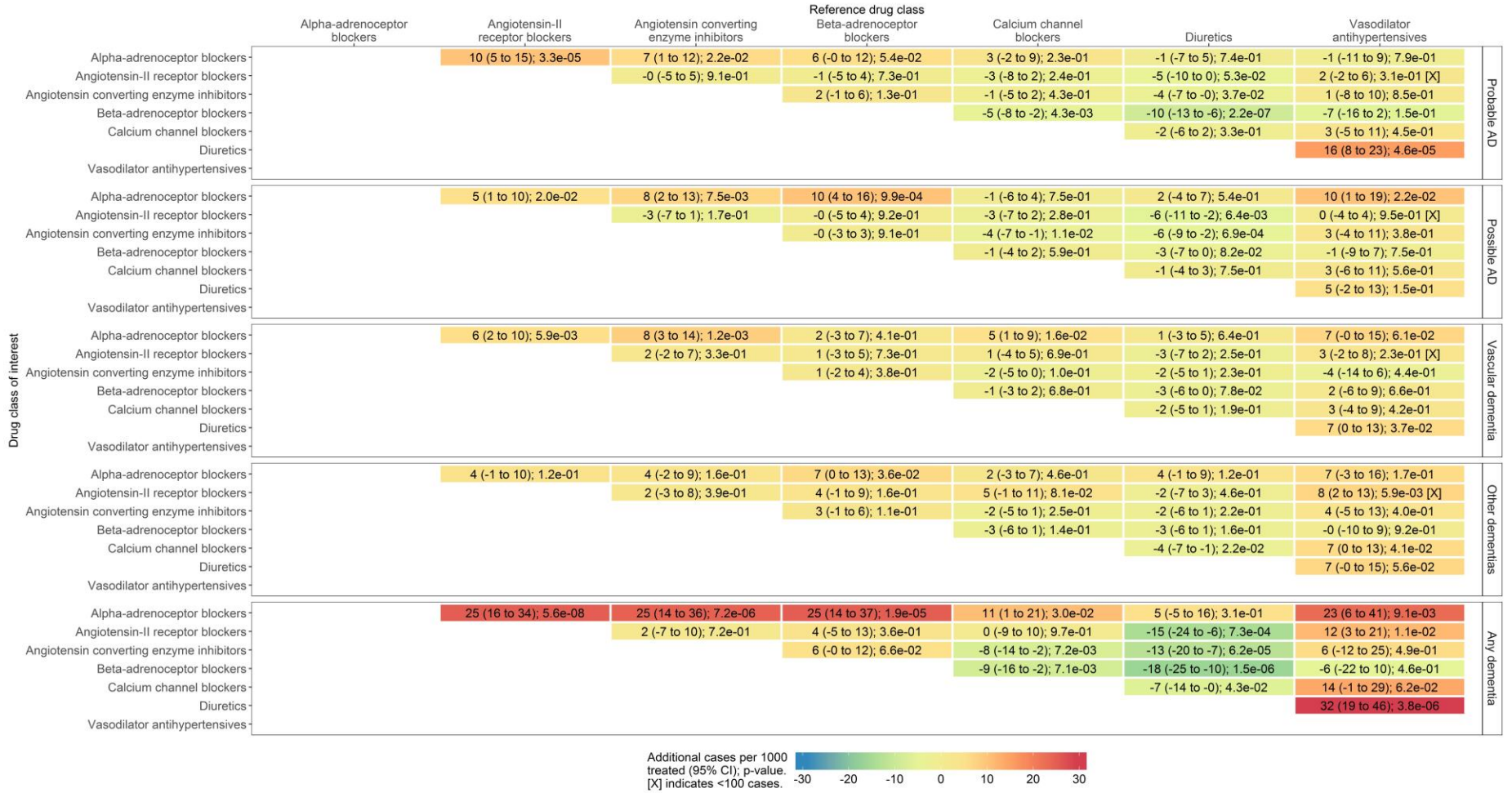
Hypertension	Centrally acting antihypertensives	Clonidine hydrochloride	*clonidine*
Hypertension	Centrally acting antihypertensives	Methyldopa	*methyldopa*
Hypertension	Centrally acting antihypertensives	Moxonidine	*moxonidine*
Hypertension	Angiotensin-II receptor antagonists	Azilsartan Medoxomil	*azilsartan*
Hypertension	Angiotensin-II receptor antagonists	Candesartan cilexetil	*candesartan*
Hypertension	Angiotensin-II receptor antagonists	Eprosartan	*eprosartan*
Hypertension	Angiotensin-II receptor antagonists	Irbesartan	*irbesartan*
Hypertension	Angiotensin-II receptor antagonists	Losartan potassium	*losartan*
Hypertension	Angiotensin-II receptor antagonists	Olmesartan medoxomil	*olmesartan*
Hypertension	Angiotensin-II receptor antagonists	Telmisartan	*telmisartan*
Hypertension	Angiotensin-II receptor antagonists	Valsartan	*valsartan*
Hypertension	Vasodilator antihypertensive drugs	Ambrisentan	*ambrisentan*
Hypertension	Vasodilator antihypertensive drugs	Bosentan	*bosentan*
Hypertension	Vasodilator antihypertensive drugs	Diazoxide	*diazoxide*
Hypertension	Vasodilator antihypertensive drugs	Hydralazine hydrochloride	*hydralazine*
Hypertension	Vasodilator antihypertensive drugs	Iloprost	*iloprost*
Hypertension	Vasodilator antihypertensive drugs	Minoxidil	*minoxidil*
Hypertension	Vasodilator antihypertensive drugs	Sildenafil	*sildenafil*
Hypertension	Vasodilator antihypertensive drugs	Sitaxentan sodium	*sitaxentan*
Hypertension	Vasodilator antihypertensive drugs	Sodium nitroprusside	*nitroprusside*
Hypertension	Vasodilator antihypertensive drugs	Tadalafil	*tadalafil*
Hypertension	PSDs and aldosterone antagonists	Amiloride hydrochloride	*amiloride*
Hypertension	PSDs and aldosterone antagonists	Eplerenone	*eplerenone*
Hypertension	PSDs and aldosterone antagonists	Spironolactone	*spironolactone*
Hypertension	PSDs and aldosterone antagonists	Triamterene	*triamterene*
Hypercholesterolaemia	Statins	Atorvastatin	*atorvastatin*
Hypercholesterolaemia	Statins	Fluvastatin	*fluvastatin*

Hypercholesterolaemia	Statins	Pravastatin sodium	*pravastatin*
Hypercholesterolaemia	Statins	Rosuvastatin	*rosuvastatin*
Hypercholesterolaemia	Statins	Simvastatin	*simvastatin*
Hypercholesterolaemia	Fibrates	Bezafibrate	*bezafibrate*
Hypercholesterolaemia	Fibrates	Ciprofibrate	*ciprofibrate*
Hypercholesterolaemia	Fibrates	Fenofibrate	*fenofibrate*
Hypercholesterolaemia	Fibrates	Gemfibrozil	*gemfibrozil*
Hypercholesterolaemia	Bile acid sequestrants	Colesevelam (hydrochloride)	*colesevelam*
Hypercholesterolaemia	Bile acid sequestrants	Colestipol (hydrochloride)	*colestipol*
Hypercholesterolaemia	Bile acid sequestrants	Colestyramine	*colestyramine*
Hypercholesterolaemia	Omega-3 fatty acid compounds	Omega-3 Acid Marine Triglycerides	*fish oil*
Hypercholesterolaemia	Omega-3 fatty acid compounds	Omega-3 Acid Ethyl Esters	*ethyl esters*
Hypercholesterolaemia	Ezetimibe	Ezetimibe	*ezetimibe*
Hypercholesterolaemia	Nicotinic acid group	Acipimox	*acipimox*
Hypercholesterolaemia	Nicotinic acid group	Nicotinic acid	*nicotinic*
Type 2 diabetes	Biguanides	Metformin (hydrochloride)	*metformin*
Type 2 diabetes	Sulphonylureas	Chlorpropamide	*chlorpropamide*
Type 2 diabetes	Sulphonylureas	Glibenclamide	*glibenclamide*
Type 2 diabetes	Sulphonylureas	Gliclazide	*gliclazide*
Type 2 diabetes	Sulphonylureas	Glimepiride	*glimepiride*
Type 2 diabetes	Sulphonylureas	Tolbutamide	*tolbutamide*
Type 2 diabetes	Other Antidiabetic Drugs	Acarbose	*acarbose*
Type 2 diabetes	Other Antidiabetic Drugs	Alogliptin	*alogliptin*
Type 2 diabetes	Other Antidiabetic Drugs	Canagliflozin	*canagliflozin*
Type 2 diabetes	Other Antidiabetic Drugs	Dapagliflozin	*dapagliflozin*
Type 2 diabetes	Other Antidiabetic Drugs	Empagliflozin	*empagliflozin*
Type 2 diabetes	Other Antidiabetic Drugs	Exenatide	*exenatide*

Type 2 diabetes	Other Antidiabetic Drugs	Linagliptin	*linagliptin*
Type 2 diabetes	Other Antidiabetic Drugs	Liraglutide	*liraglutide*
Type 2 diabetes	Other Antidiabetic Drugs	Nateglinide	*nateglinide*
Type 2 diabetes	Other Antidiabetic Drugs	Pioglitazone	*pioglitazone*
Type 2 diabetes	Other Antidiabetic Drugs	Repaglinide	*repaglinide*
Type 2 diabetes	Other Antidiabetic Drugs	Rosiglitazone	*rosiglitazone*
Type 2 diabetes	Other Antidiabetic Drugs	Saxagliptin	*saxagliptin*
Type 2 diabetes	Other Antidiabetic Drugs	Sitagliptin	*sitagliptin*
Type 2 diabetes	Other Antidiabetic Drugs	Vildaglipin	*vildagliptin*
Type 2 diabetes	Insulin products	Insulin	*insulin*

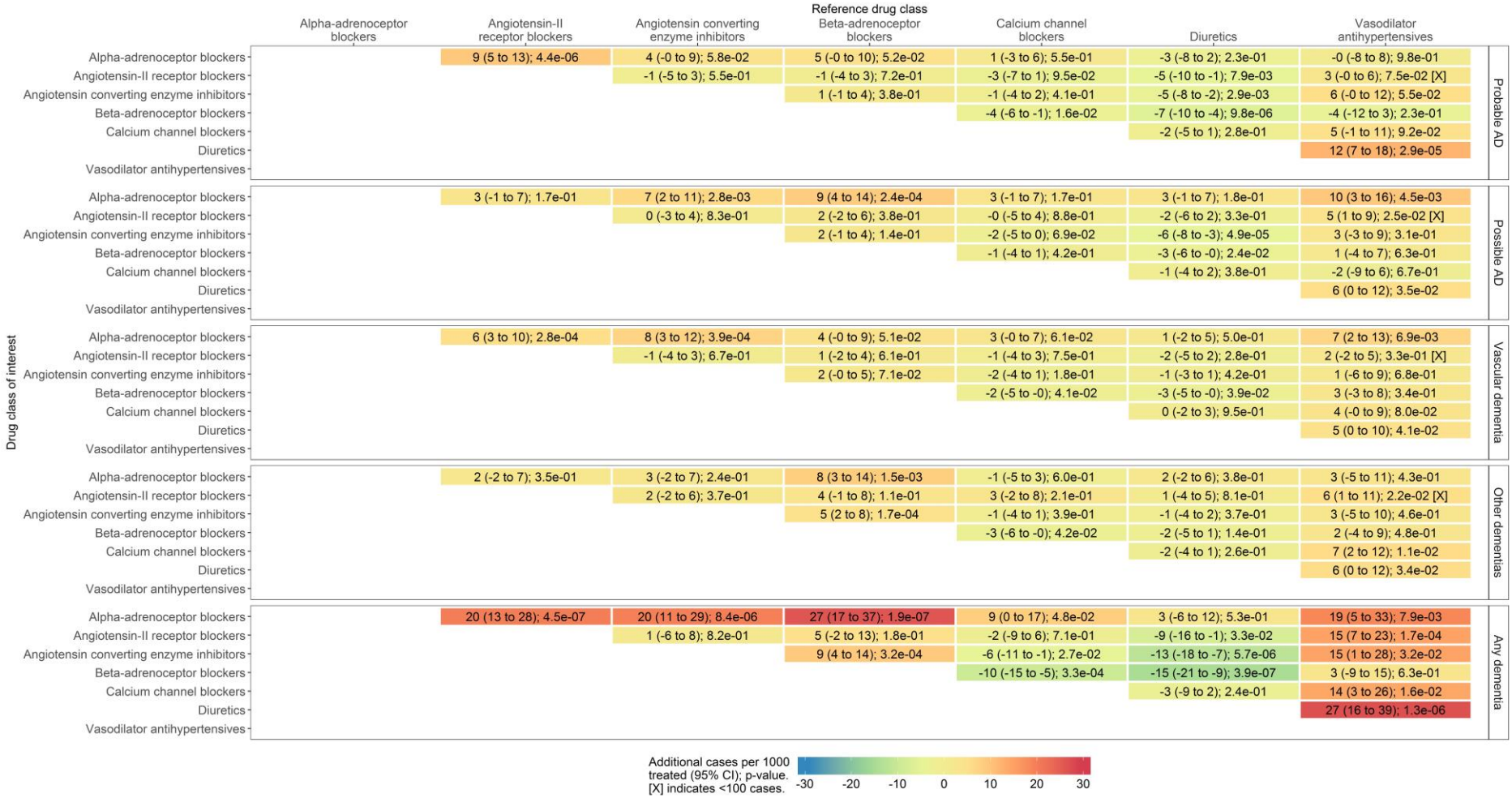
Appendix C. Additional sensitivity analyses for the assessment of antihypertensive drugs using electronic health record data

Figure B.1: Instrumental variable estimates for all dementia outcomes after adjustment for socioeconomic position using electronic health record data.



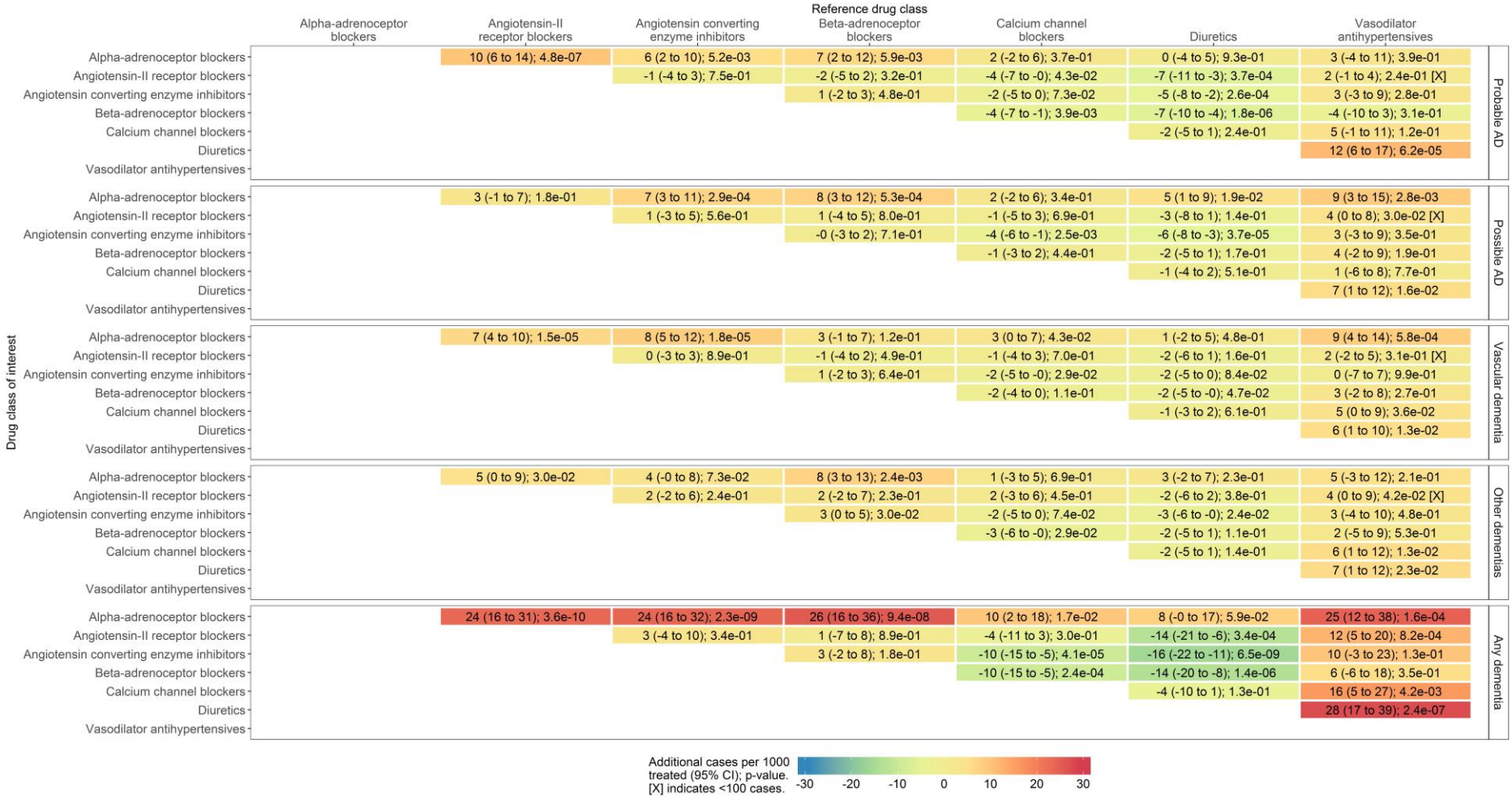
F greater than 2530 for all analyses.

Figure B.2: Instrumental variable estimates for all dementia outcomes after adjustment for body mass index using electronic health record data.



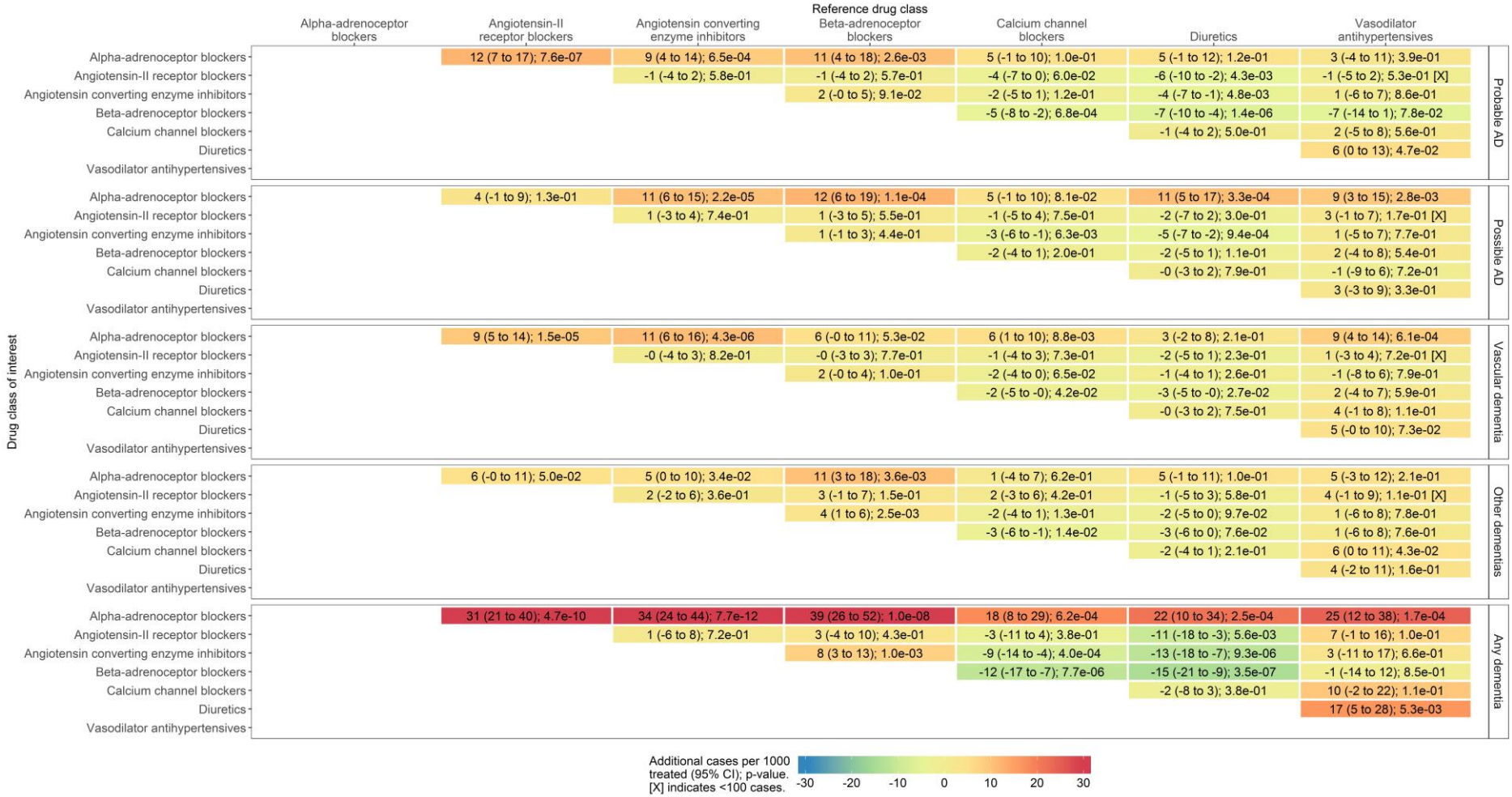
F greater than 3890 for all analyses.

Figure B.3: Instrumental variable estimates for all dementia outcomes after adjustment for chronic disease using electronic health record data.



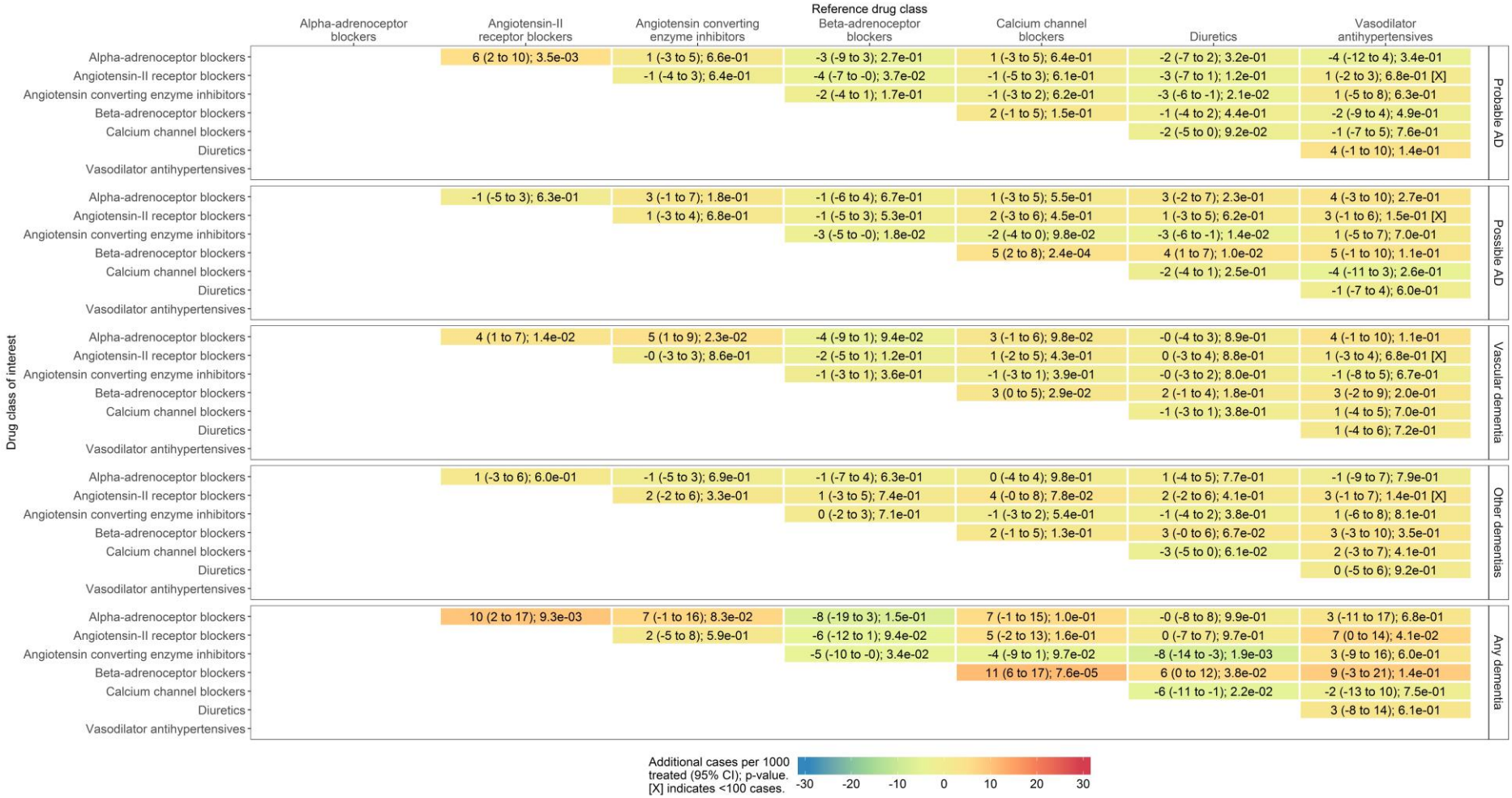
F greater than 4499 for all analyses.

Figure B.4: Instrumental variable estimates for all dementia outcomes after adjustment for sex using electronic health record data.



F greater than 4416 for all analyses.

Figure B.5: Instrumental variable estimates for all dementia outcomes after adjustment for age using electronic health record data.



F greater than 4094 for all analyses.