Using physician's prescribing preference as an instrumental variable in comparative effectiveness research

Lisong Zhang

BSC, MSC

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

Health economics and health technology assessment

School of Health and Wellbeing

College of Medical, Veterinary & Life Sciences

University of Glasgow

December 2023

Abstract

Background

Comparative effectiveness research (CER) studies using non-randomised study designs sometimes employ instrumental variables (IVs) to address the problem of unmeasured confounding. Physician's prescribing preference (PPP) is a commonly used IV in this context and had been shown to have utility in many CERs. However, these IVs are generally used as a supplementary method rather than the main analytical strategy. In this thesis, I aim to test the validity of PPP IVs, including an evaluation of the different ways they can be constructed to help promote their more widespread use in CER.

Methods

This thesis consists of a range of underpinning methodological approaches, including a literature review summarising applied and simulation studies between 2005 and 2020 that use PPP IV in CER; applied CERs using PPP IV in studies utilising routinely-collected health datasets; target trial emulation approaches based on benchmarking from a randomised clinical trial; and simulation studies to test the performance of PPP IV in multiple CER settings.

Results

My literature review provides guidance on the further use of physician's prescribing preference as instrumental variables in comparative effectiveness research. It highlighted that practical use of PPP needs to consider the findings from simulation studies in the area. In my empirical chapters, I provide strong evidence that PPP is a valid IV approach for conducting CERs using non-randomised study designs. I found that constructing PPP using longer prescription histories generally produces stronger instruments, which in turn leads to greater precision in estimation of treatment effects. In practice, validation of assumptions is crucial for the utility of IVs in CER. In my applied research, I found strong real-world evidence that supports diazepam is associated with lower risk of rehospitalisation and mortality due to the alcohol intoxication and harmful than chlordiazepoxide; that disulfiram is superior to acamprosate in terms of preventing alcohol dependence-related hospitalisations; and that sulfonylureas (SU) performs better than dipeptidyl peptidase-4 inhibitor (DPP-4 inhibitor) in reducing HbA1c levels as the second-line treatment for Type-2 diabetes patients. In my simulation studies, I found PPP IV, when unmeasured confounding exists, can produce less biased estimates of treatment effects than conventional multivariable regressions that only adjust for measured confounding variables, albeit with lower statistical power. The

simulations also show PPP IV has potential in alleviating noncollapsibility in non-linear IV approaches.

Implications

Findings from this thesis indicate that PPP IVs can be valid IVs and reduce unmeasured confounding in observational CER studies. However, I have found that there is room for improvement in the application of PPP IV in CER studies; researchers need to pay more attention on validating IV assumptions and carefully consider how different formulations of PPP IVs can be applied in order to improve the quality of statistical inference. Future applied PPP IV research should consider findings from relevant simulation studies to inform study designs and analysis plans. Conversely, one also needs information on PPP IVs from empirical studies to inform future simulation study design and to gain further knowledge from triangulation between applied and simulation findings. Many of my thesis findings can be generalised to the use of non-PPP IV approaches in CER.

Table of contents

Abstract 1

List of tables 9

List of figures 10

Abbreviation 11

Acknowledgement 12

Author's Declaration 13

Contribution statement 14

Competing interests 15

Chapter 1. General introduction 16

1.1 Comparative effectiveness research in observational studies 16

1.2 The unmeasured confounding issue in the observational studies 17

1.3 Introduction of the use of instrumental variable in CER 18

1.4 Introduction of the use of physician's prescribing preference (PPP) as instrumental variable and other types of preference-based instrumental variable 19

1.5 Structure of this thesis 19

1.6 Aims and objectives 20

Chapter 2. Literature review: An introduction and commentary on the implementation of PPP IV in comparative effectiveness research 21

2.1 Introduction 21

2.2 The validation of IV assumptions 22

2.2.1 Validation of the relevance assumption 25

2.2.2 Validation of exclusion restriction 26

2.2.3 Validation of independence assumption 27

2.2.4 Validation of treatment homogeneity and monotonicity assumption 27

2.2.5 Specific PPP IV considerations 28

2.3 Type of treatment effects that IV method estimates 30

2.4 Estimation methods for IV method 30

2.4.1 Specific PPP IV considerations 33

2.5 Discussion....33
2.6 Conclusion....34
Chapter 3. Physician's prescribing preference instrumental variables: comparative effectiveness research should consider methodological insights from simulation studies. 35
3.1 Authorship, and publication details....35
3.2 Abstract....36
3.3 Introduction 38
3.4 Method 39
3.4.1 Inclusion and Exclusion Criteria 39
3.4.2 Search terms 39
3.5 Results....39
3.5.1 Summary of the simulation studies 41
3.5.2 Summary of the applied studies 42
3.6 Discussion....46
3.7 Conclusion....49
Chapter 4. Using Physician's prescribing preference as an instrumental variable to compare the effectiveness of diazepam and chlordiazepoxide hydrochloride for patients diagnosed with alcohol intoxication and harmful use. 50
4.1 Publication details....50
2.3.1 Specific for PPP IV considerations....50
4.2 Data used in this chapter. 50
4.3 Abstract....53
4.4 Introduction 55
4.4.1 Background 55
4.4.2 Pharmacological treatment of AUD (alcohol withdrawal syndrome)....55
4.4.3 Pharmacological treatment of AUD (with history of alcohol intoxication and harmful use) 56
4.5 Method 58
4.5.1 Physician's prescribing preference as an instrumental variable....58

4.5.2 Validation of assumptions....60
4.5.3 Study population....61
4.5.4 Statistical methods....62
4.5.5 Sensitivity analysis for unmeasured confounding....63
4.6 Results....64
4.6.1 Descriptive statistics....64
4.6.2 Validation of IV assumptions....66
4.6.3 Comparison of estimates of the treatment effect....68
4.6.4 Survival analysis....70
4.7 Discussion....72
4.8 Strengths and Limitations....75
4.9 Conclusion....76
Chapter 5. Epidemiology of pharmacological treatments for alcohol dependence in the UK: evidence from primary and secondary healthcare data....77
5.1 Publication and author contribution details....77
5.2 Data used in this chapter.....77
5.3 Abstract....79
5.4 Introduction....80
5.5 Methods....83
5.5.1 Data sources....83
5.5.2 Pharmacological treatments....83
5.5.3 Statistical analyses....84
5.6 Results....88
5.6.1 Trends and rates....88
5.6.2 Inequality....89
5.6.3 Comparative effectiveness....92
5.6.4 Sensitivity analysis for the unmeasured confounding....93
5.7 Discussion....93

5.8 Strengths and Limitations 98

5.9 Conclusion 100

Chapter 6. Comparing DPP-4 inhibitor and sulfonylurea as second-line treatment for T2DM patients: a target trial emulation 101

6.1 Publication details 101

6.2 Data used in this chapter. 101

6.3 Abstract 103

6.4 Introduction 104

6.5 Approach 1 106

6.5.1 Method 106

6.5.2 Results 111

6.6 Approach 2 116

6.6.1 Method 116

6.6.2 Results 116

6.7 Discussion 119

6.8 Strengths and Limitations 122

6.9 Conclusion 122

Chapter 7. Assessing the performance of Physician's Prescribing Preference as an instrumental variable in Comparative Effectiveness Research with moderate and small sample sizes: a simulation study 123

7.1 Publication details 123

7.2 Abstract 124

7.3 Introduction 125

7.4 Method 126

7.4.1 Statistical analysis approaches 126

7.4.2 Simulation design 127

7.5 Results 129

7.6 Discussion 135

7.7 Conclusion 136

Chapter 8. Comparing the performance of two-stage prediction substitution and two-stage residual inclusion methods when using physician's prescribing preference as an instrumental variable in comparative effectiveness research. 137

8.1 Publication details 137

8.2 Abstract 137

8.3 Introduction 139

8.4 Method 140

8.4.1 Data generating process. 140

8.4.2 Study design 142

8.4.3 Design of the empirical illustration 143

8.5 Results 144

8.5.1 Research objective 1: Assessing the ability of 2SRI and 2SPS of alleviating the unmeasured confounding bias. 144

8.5.2 Research objective 2: Assessing the ability of 2SPS and 2SRI of alleviating noncollapsibility. 146

8.7 Strengths and Limitations 150

8.6 Discussion 150

8.8 Conclusion 152

Chapter 9. General discussion 153

9.1 Research objectives of this thesis 153

9.2 Summary of thesis chapters 153

9.2.1 Part 1 154

9.2.2 Part 2 154

9.2.3 Part 3 155

9.3 Gaps identified in this thesis. 155

9.4 Strengths and Limitations 158

9.5 Implications for further research 159

9.6 Overall conclusion of the thesis 162

List of references 163

Supplementary material 177

List of tables in supplementary material 177

List of figures in supplementary material 177

Supplementary material for Chapter 3 178

Supplementary material for Chapter 4 179

Data request form for Chapter 4 183

Supplementary material for Chapter 5 186

Data request form for Chapter 5 (CPRD) 193

Supplementary material for Chapter 6 210

Data request form for Chapter 6 212

GitHub link for the R code used in this thesis 219

List of tables

Table 1. Table of notation 23
Table 2. Notation of compliance classes for a binary IV 24
Table 3. Eight combinations of Y, Z, D 24
Table 4. Summary of common IV estimation methods 32
Table 5. Summary of estimation methods in different settings 43
Table 6. Validation of assumptions 45
Table 7. Checklist for the researchers who are interested in using PPP in CERs 49
Table 8. Descriptive statistics of the study population 65
Table 9. Validation of the relevance assumption using F-statistics and AUC. 66
Table 10. Characteristics of cohorts used in inequality and comparative effectiveness analyses. 87
Table 11. Inequality models of prescriptions after hospitalisation, before hospitalisation and inequality in prescriptions between the two most used alcohol dependence drugs 90
Table 12. Protocol for target emulation trial 107
Table 13. Descriptive statistics of the study population 112
Table 14. OLS, propensity score and 2SLS estimates. 113
Table 15. Average treatment effect and local treatment effect estimated by MTE methods. 95% CI was bootstrapped by 1000 times. 114
Table 16. Estimated OR from logistic regression. 117
Table 17. 2SRI estimates of binary outcomes. 118
Table 18. Measurement of performance. 126
Table 19. Measurement of performance 142
Table 20. Results from empirical case study. β' represents the estimate from unadjusted logistic regression. β represents the estimate from adjusted logistic regression. $\beta IPSW$ represents the estimate from the IPSW using stabilised weight. $\beta 2SRI$ represents the estimate from adjusted 2SRI. 148
Table 21. Gaps identified in this thesis. 157
Table 22. Guidance and key considerations for the further research 161

List of figures

Figure 1. PPP is the unmeasured instrument U*. Z is the proxy of U*. X is the actual treatment assigned and Y is the outcome, U represents unmeasured confounding. Arrows indicate causal relationships. The directions of the arrows represent the direction of the causality. 22
Figure 2. Flow chart of literature search 40
Figure 3. Construction of PPP IV 59
Figure 4. OLS, propensity score and 2SLS estimates. 69
Figure 5. Conventional Cox regression and IV-based Cox regression results 70
Figure 6. Sensitivity analysis results for death as outcome, the blue contour shows that sensitivity parameters corresponding estimated treatment effect. 71
Figure 7. Trends and rates of alcohol dependence diagnosis (a) and alcohol use disorders hospitalisation (b) and percentage of such individuals receiving alcohol dependence prescriptions within 60 days of their diagnosis (a) or discharge date (b). In panel b 2019 was removed as data were only until March. 88
Figure 8. Point estimate of hazards ratio for AUD hospitalisation and confidence intervals of the models measuring comparative effectiveness research of disulfiram and acamprosate (using acamprosate as reference). 92
Figure 9. Flow chart of defining study cohort 108
Figure 10. Percent bias of 2SLS and OLS. 129
Figure 11. Coverage rate across 1000 times simulation. The blue intercept line represents the nominal 95% 131
Figure 12. p values of OLS and 2SLS estimates. 132
Figure 13. 95% Confidence intervals of OLS and 2SLS estimates (95% CIs are calculated using cluster robust standard errors). The red line represents the null hypothesis. 134
Figure 14. Estimates from GLM, 2SRI and 2SPS. 144
Figure 15. Coverage rates from 2SRI, 2SPS, GLM. Red dash line represents the 95% nominal. 145
Figure 16. Percent bias of 2SRI, 2SPS, GLM when there is noncollapsibility. Percent bias (a) represents percent bias of the estimate when $\gamma 2$ equals 1.0. Percent bias (b) represents percent bias of the estimate when $\gamma 2$ equals 1.5. 146
Figure 17. Coverage rate of GLM, 2SRI and 2SPS. Coverage rate (a) represents coverage rate of the estimate when $\gamma 2$ equals 1.0. Coverage rate (b) represents coverage rate of the estimate when $\gamma 2$ equals 1.5. 147
Figure 18. Published CER using PPP as IV from 2005 to 2023 160

Abbreviation

2SLS	Two-stage Least Square
2SPS	Two-stage Prediction Substitution
2SRI	Two-stage Residual Inclusion
AIH	Alcohol Intoxication and Harmful Use
ATE	Average Treatment Effect
ATT	Average Treatment Effect On The Treated
AUD	Alcohol Use Disorder
ATU	Average Treatment Effect On The Untreated
AWS	Alcohol Withdrawal Syndrome
BVP	Bivariate Probit Model
CACE	Conditional Average Treatment Effect
CART	Classification and Regression Tree
CCI	Charlson Comorbidity Index
CER	Comparative Effectiveness Research
CPRD	Clinical Practice Research Datalink
DPP4-inhibitor	Dipeptidyl peptidase-4 inhibitor
GLM	Generalised Linear Model
GMM	Generalised Methods of Moments
HR	Hazards Ratio
ICD-10	International Classification of Diseases 10th Revision
IMD	Index of Multiple Deprivation
IV	Instrumental Variable
LATE	Local Average Treatment Effect
LIV	Local Instrumental Variable
MTE	Marginal Treatment Effect
OR	Odds Ratio
PIS	Prescribing Information System
PPP	Physician's Prescribing Preference
SDRN	Scottish Diabetes Research Network
SIMD	Scottish Index of Multiple Deprivation
SMR	Scottish Morbidity Record
SMM	Structural Mean Model
SU	Sulfonylureas
T2DM	Type 2 Diabetes Mellitus
UK	United Kingdom
WHO	World Health Organisation

Acknowledgement

I would express my deepest appreciation to my supervisors Prof. Jim Lewsey and Prof. David McAllister for their guidance. I am also grateful to my colleague Mr. Francesco Manca, with whom I collaborated on one chapter of this thesis (Chapter 5), and other works outside of this thesis.

Special thanks to Ms. Johanna Bruce and Ms. Judi Evans from Scottish Safe Haven, and Dr. Luke Blackbourn from University of Edinburgh. Without their support, I could have not completed my data analysis of this thesis.

Last but not least, I would express my gratitude to my family and friends. During three-year pandemic, I could not make it without the remote support from them. Backing home after spending 1232 days alone in Glasgow, they comforted me and helped me through the most stressful phase of completing this thesis.

After this long, hard, and fruitful journey, I also need to say Thank you to myself. Hereby I translated a poem from Li Shangyin from Tang dynasty which depicts a pine tree.

Here in this yard, you grow alone.

Slim leaves bring shade, and wind blown.

You feel lost when the plum blossom blooms.

After snow, the greenness makes you the lasting tone.

Author's Declaration

Hereby I declare that this work has not been submitted in any form for another degree at the University of Glasgow or other institutions. I am the sole author of this thesis, excluding the contribution of others that has been acknowledged as below.

Contribution statement

Chapter 3

I conceptualised and developed the research question for this literature review. I searched the literature, critically analysed the results, and drafted the original manuscript. Jim Lewsey and David McAllister critically reviewed and revised the draft.

Chapter 5

I worked with Francesco Manca as the joint first author, responsible for the conception and design of the study, data analysis, visualisation, data interpretation under Jim Lewsey's supervision. Francesco Manca produced the original draft. I, Francesco Manca, Jim Lewsey, Claire Sharp, Niamh Fitzgerald, Andrew McAuley, Hamish Innes, Vittal Katikireddi, Frederick Ho, Bhautesh Jani, critically revised and commented the final version of the manuscript.

Chapter 6

This chapter was based on the PERMIT project led by Professor Richard Grieve in London School of Health and Tropical Medicine. The research question of this chapter is originally from the PERMIT project. I led the application of PPP IV using data from SDRN. Jim Lewsey and David McAllister critically reviewed and revised the manuscript.

Chapter 7

I conceptualised and developed the research question, designed and analysed the simulation study, and drafted the manuscript. Jim Lewsey and David McAllister critically reviewed and revised the manuscript.

Chapter 8

I conceptualised and developed the research question, designed the simulation study, analysed the results and drafted the original manuscript. Jim Lewsey and David McAllister critically reviewed and revised the manuscript.

Competing interests

I declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

In addition to the chapters outlined above, I also collaborated and published with HEHTA colleagues on research into evaluating the effect of Scotland's minimum unit pricing for alcohol intervention on alcohol dependence outcomes [Manca, F., Zhang, L., Fitzgerald, N. *et al.* The Effect of Minimum Unit Pricing for Alcohol on Prescriptions for Treatment of Alcohol Dependence: A Controlled Interrupted Time Series Analysis. *Int J Ment Health Addiction* (2023). https://doi.org/10.1007/s11469-023-01070-6.]

Chapter 1. General introduction

1.1 Comparative effectiveness research in observational studies

To make well-informed decisions in the health sector and improve patients' health, comparative effectiveness research (CER) is used in comparing the effectiveness and the risks of therapies or medical interventions (National records of Scotland). In CER, randomised controlled trials (RCTs) are considered the gold standard in terms of causal inference (Greenland, 2000). It is mainly because RCTs randomise the treatment allocation which ensures balance in the patients' characteristics in treated and controlled groups. However, RCTs are not always feasible. Ethical and financial problems are common issues. Also, possible non-adherence and loss of follow-up make treated and controlled group less comparable. Although the intention-to-treat (ITT) estimator is effective in reducing non-compliance bias in RCTs, the substantial lack of adherence makes ITT estimates less likely to provide a consistent estimate of treatment effectiveness (Hernán and Hernández-Díaz, 2012). The strict inclusion and exclusion criteria of RCTs ensure the internal validity but it also make the study cohort less representative (Booth and Tannock, 2014). Besides, it only evaluates the efficacy of treatments (how a treatment works under the ideal situation) but not effectiveness (how a treatment works in real-world) (Faraoni and Schaefer, 2016, Nallamothu et al., 2008). In order to get unbiased estimates of effectiveness, one can consider using observational studies which use data from population-based routinely-collected datasets (Faraoni and Schaefer, 2016).

As another type of CER, observational studies do not have pre-defined intervention of interest therefore is able to investigate multiple types of interventions (Sørensen et al., 2006). The volume of participants in the observational data is usually large which is more likely to generalised to whole population resulting strong external validity. This is particularly important for vulnerable group and rare conditions which is not feasible or costly to study in RCTs (Armstrong, 2012). Also, it makes investigation of chronic conditions possible as the data can cover a long time span (Booth and Tannock, 2014).

1.2 The unmeasured confounding issue in the observational studies

Observational study has limits. Since observational studies usually involve data that are already been collected, it is sometimes not possible to compare the effectiveness of new treatments as the data either does not exist or being immature (Armstrong, 2012). More importantly, observational study is subject to bias, arguably, concerns about bias are larger for observational than randomised studies. The information bias, which is introduced by measurement error or misclassification can affect the casual inference in RCTs as well as in observational studies. In observational study settings, data are collected for the purpose which may not be consistent with the research objective. Therefore, the measurement of treatment and outcome variables are not optimal which leads to bias in the treatment effect estimation (Faraoni and Schaefer, 2016). Selection bias is another source of bias. It occurs when conditioning on colliders or common effects of treatment and outcome. In this case, the selection mechanisms are not random. Weighting adjustment methods can be used to alleviate selection bias in surveys and cohort analysis (Greenacre, 2016), such as propensity score and inverse probability-of-censoring weighted estimation (Howe et al., 2016, Thompson and Arah, 2014).

The confounders are the common causes of treatments and outcomes. If these variables cannot be controlled or measured or recorded, causal effect between exposure and treatment is introduced in error which leads to confounding bias. Confounding bias and selection bias can happen simultaneously (Haneuse, 2016). However, unlike selection bias, unmeasured confounding bias is more difficult to address. Multivariable regression models and propensity score are commonly used analytical tools to control the confounding bias in observational studies (Austin, 2011). However, some studies pointed out the drawbacks of these ‘standard’ methods. In a review paper that summarises six systematic reviews on the comparison on the of treatment effect estimates from propensity score matching and RCT, the disagreement between the RCT and propensity score happened in 68 of 127 (54%) comparisons (measured by ratio of relative risk from RCT over the relative risk from propensity score larger than 1.43 or less than 0.7) (Forbes and Dahabreh, 2020). Besides, same as regression adjustment, propensity score can only control measured confounding issue (Ali et al., 2015). The instrumental variable method is designed to bypass this issue.

1.3 Introduction of the use of instrumental variable in CER

Originating from econometrics, instrumental variable (IV) has been widely applied in CER using observational data in recent decades (Hernán and Robins, 2006). One of most used IV is preference-based IV (Brookhart et al., 2006). Preference-based IV assumes that the preference of the treatment provider can strongly affect the treatment assignment. Since it is a natural process, it is unlikely be associated with the outcome and the other covariates (Brookhart and Schneeweiss, 2007).

There are subtypes of facility-level prescribing preference (e.g., hospital-level prescribing preference) and individual-level prescribing preference. For individual-level prescribing preference, such as physician's prescribing preference (PPP), if the prescribing preference is strong and homogeneous over a fixed period of time, then PPP can potentially be a valid IV. There are studies proved that the hospital-level preference performs better than individual-level preference in terms of reducing the unmeasured bias (Ionescu-Ittu et al., 2012). However, the facility-level preference is more difficult to maintain at a stable pattern compared with the individual-level (Potter et al., 2020).

Also exploiting the concept of 'naturally variation', calendar time and distance to the facility are also commonly used IV in observational CERs (Chen and Briesacher, 2011, Ertefaie et al., 2017). In terms of investigating acute conditions, the distance to facility is highly related to the treatment received. However, the distance IVs in these cases tend to have direct effect on the outcome which violate the exclusion restriction assumption (Baiocchi et al., 2014) . Calendar time can be considered as an IV when the policy or guideline changes have direct impact on the choice of treatment. Nevertheless, time variables may also be associated with the unmeasured cofounders as the time may bring changes in characteristics in the participants that enter the cohort (Brookhart et al., 2010).

This thesis will focus on assessing the performance of PPP as instrumental variable in CER using empirical as well as simulation studies.

1.4 Introduction of the use of physician's prescribing preference (PPP) as instrumental variable and other types of preference-based instrumental variable

Brookhart and colleagues firstly proposed PPP as a potential valid IV (Brookhart et al., 2006). As it has natural-occurred variation which make the treatment assignment close to that in RCT, PPP IV has been increasingly used in observational studies (Brookhart et al., 2010, Boef et al., 2016a, Ionescu-Ittu et al., 2012). Since it is a latent variable that cannot be measure directly, different forms of surrogate variables are proposed. The most widely used one is prescribing history of the physicians with an intuition that the prescribing preference can be reflected from past prescribing behaviour (Boef et al., 2016b).They used the previous prescriptions of one particular physician as the proxy for the prescribing preference of that physician (Davies et al., 2013c, Davies et al., 2013a, Brookhart and Schneeweiss, 2007). Examples include the most recent prescription and several prior prescriptions or the proportion of one drug (Kollhorst et al., 2016). Like the PPP, other forms of preference-based IV also assume that providers have different preference for different types of treatment which have direct effect on the treatment assignment (Brookhart and Schneeweiss, 2007).

1.5 Structure of this thesis

This thesis is organised as follows:

Chapter 2 contains a general review of IV approach and a critical commentary of using PPP as IV in CER. Chapter 3 is a literature review contains two parts: 1) review of CERs that use routinely-collected data; 2) review of the simulation studies that build hypothetical drug comparison studies. The aim is to review the methodology used in current CERs that investigate the PPP IV and identify possible methodological advances that simulation studies proposed. Chapter 4, Chapter 5 and Chapter 6 are three CERs using three data sources: Scottish National Prescribing Information System (PIS), Clinical Practice Research Data (CPRD) and Scottish Diabetes Research Network (SDRN) respectively. The aim of these chapters is to assess the performance of different forms of proxy for PPP. Chapter 7 and Chapter 8 are simulation studies that investigate the performance of IV approaches in

different scenarios. I aim to assess the utility of PPP IV in different research settings. Chapter 9 is the general discussion and conclusion.

1.6 Aims and objectives

This thesis has three major aims:

1. Critically review the current comparative effectiveness research which use PPP as IV
2. Implement PPP IV in drug comparison studies using routinely collected data.
3. Explore the novel use of PPP IV in different settings, using real-world data as well as simulated data.

Chapter 2. Literature review: An introduction and commentary on the implementation of PPP IV in comparative effectiveness research

2.1 Introduction

For CER, RCT is acknowledged to be the gold standard methodology. It allows the characteristics of patients in treatment and control group to be balanced, on average, therefore eliminating the confounding issue (Greenland, 1990). However, RCTs are not always feasible. Observational studies that utilise routinely-collected dataset are widely applied in CERs (Armstrong, 2012). Observational data do not usually record sufficiently the potentially confounding variables which affect both the treatment assignment and the outcome(s). Such unrecorded variables, and other confounding variables that are unknown or not anticipated, lead to unmeasured confounding bias. As an analytical method, IV method is mainly used to alleviate unmeasured confounding bias.

There are a number of articles that provide a general introduction of the IV method in clinical epidemiology literature in recent years (Chen and Briesacher, 2011, Baiocchi et al., 2014, Lousdal, 2018, Potter et al., 2020, Widding-Havneraas and Zachrisson, 2022). However, most of this literature focuses on the general form of IVs rather than PPP IV. PPP is a commonly used IV in CER (Brookhart et al., 2006, Davies et al., 2013c, Davies et al., 2020). As a latent variable, PPP IV cannot be measured directly and requires a proxy of the 'true' prescribing preference (see Figure 1).

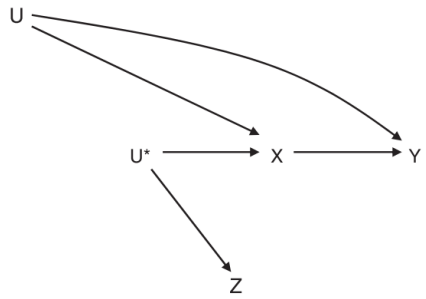


Figure 1. PPP is the unmeasured instrument U. Z is the proxy of U*. X is the actual treatment assigned and Y is the outcome, U represents unmeasured confounding. Arrows indicate causal relationships. The directions of the arrows represent the direction of the causality.*

As it mentioned in Chapter 1, PPP exploits the natural variation to be a substitute for the 'flipping coin' treatment assignment mechanism in RCTs. Despite the advantage from the natural variation in PPP, the implementation of PPP IVs brings extra important considerations which have not been fully critically evaluated in the literature to date. Therefore, the objective of this chapter is to provide an introduction and commentary of PPP IVs to assist researchers conducting CER with PPP IV. This chapter is arranged in three parts: 1) validation of IV assumptions; 2) treatment effects that IV method estimates; 3) estimation methods for IV estimates. In each part, I first provide a general introduction of using IV in the setting of CER, then a specific introduction of PPP IVs with the purpose of providing a guidance framework for using PPP IV.

2.2 The validation of IV assumptions

Before the introduction of IV assumptions in more detail, a table of notation is presented (see Table 1).

Notation	**Meaning**
Y^1,Y^0	Counterfactual outcome if the treatment is 1 or 0
D^1, D^0	Treatment received when the IV is 1 or 0
Z	IV
L	Measured confounders
U	Unmeasured confounders
i	Participants

Table 1. Table of notation

(1) Stable Unit Treatment Value Assumption (SUTVA)

One general assumption for estimating the treatment effect is the stable unit treatment value assumption (SUTVA) which was proposed by Rubin (Rubin, 1986). It states that the treatment effect of one individual is not affected by the treatment received by others. One possible scenario that SUTVA is violated when the students in treatment group and the students in control group stay in the same classroom where they may interact (Stuart, 2010). Note that the non-compliance in RCT clearly violates SUTVA as it makes the treatment assignment different from the treatment received. Likewise, in observational studies, one cannot measure what exposure the patients actually received (Schwartz et al., 2012).

Apart from SUTVA, a valid IV should satisfy three assumptions (Hernán and Robins, 2006):

1. Relevance assumption: IV is associated with the exposure.
2. Exclusion restriction assumption (ER assumption): IV only affects the outcome through the exposure.
3. Independence assumption: IV is not associated with unmeasured confounders.

However, to get the treatment parameter effect estimate, one needs further assumptions around the relationship between the treatment received and what the instrument indicates. Leading to that, consider the notation of compliance classes presented in Table 2. In the simple case of a binary IV, there are four types of compliance classes based on participant *i*:

Compliers	$D_i^1 = 1; D_i^0 = 0;$
Never takers	$D_i^1 = 0; D_i^0 = 0;$
Always takers	$D_i^1 = 1; D_i^0 = 1;$
Defiers	$D_i^1 = 0; D_i^0 = 1;$

Table 2. Notation of compliance classes for a binary IV

Combined with a potential binary outcome, there are eight combinations that can be observed (see Table 3).

Y	**Z**	**D**	**Definition**
1	1	1	Compliers or always takers
1	1	0	Defiers or never takers
1	0	1	Defiers or always takers
1	0	0	Compliers or never takers
0	1	1	Compliers or always takers
0	1	0	Defiers or never takers
0	0	1	Defiers or always takers
0	0	0	Compliers or never takers

Table 3. Eight combinations of Y, Z, D

In order to get the parameter estimate of the treatment effect, one needs to know the eleven probabilities: Pr(Y^1=1|Complier); Pr(Y^0=1|Complier); Pr(Y^1=1|Always takers); Pr(Y^1=1|Defiers); Pr(Y^0=1|Defiers); Pr(Y^0=1|Never takers), and Pr(compliers), Pr(never takers), Pr(defiers), Pr(always takers) and Pr(Z=1). Since the sum of the proportions of the four compliance groups is 1, the degrees of freedom are reduced to 10. Still, the model is not fully defined (10 unknown parameters vs. 8 observed combinations of Y, Z, D).

One possible additional assumption to reduce the unknown parameters to estimate is to assume treatment effect homogeneity among the four compliance classes (i.e., E[$Y^1 - Y^0$| compliers]=E[$Y^1 - Y^0$| never takers]=E[$Y^1 - Y^0$| always takers]=E[$Y^1 - Y^0$| defiers]) (Baiocchi et al., 2014). As it mentioned in the Chapter 1, under this assumption, the IV method estimates ATE rather LATE. The validation of treatment homogeneity is difficult, since evidence of treatment effect heterogeneity is commonplace in epidemiology (Poole et al., 2015, Labrecque and Swanson, 2018). Since one cannot truly observe the proportion of defiers, another type of additional assumption is to assume no defiers which is also referred to as no marginal effect subjects (Harris and Remler, 1998).

Imbens and Angrist first introduced the monotonicity assumption for IV method (Imbens and Angrist, 1994). Angrist proposed that one should rule out the defier group which makes the treatment received monotonically related with the IV (Angrist et al., 1996). In never takers and always takers groups, the instrument does not affect the treatment assignment. Thus, under the monotonicity assumption, IV method only estimates the LATE which is the treatment effect in the compliers group (Hernán and Robins, 2006).

2.2.1 Validation of the relevance assumption

The relevance assumption is the only IV assumption that can be validated empirically (Hernán and Robins, 2006, Labrecque and Swanson, 2018). F-statistics and R-squared from a regression model fitted to the exposure and IV are commonly used in measuring the strength of association between the treatment group and the IV. The ‘rule of thumb’ for a strong enough instrument is that the F-statistics should be greater than 10, which is based on the critical value that makes the relative bias of the 2SLS estimator based on the OLS estimator less than 0.1 in a weak instrument test (Stock and Yogo, 2002, Staiger and Stock, 1994). When the IV is binary, one can also use odds ratios (OR) and C-statistics to measure the strength of association between the exposure and the IV (Pratt et al., 2010).

2.2.2 Validation of exclusion restriction

This assumption is not empirically verifiable. One can only use falsification tests to assess it indirectly. One straightforward approach is the IV inequality method (Balke and Pearl, 1997) which is presented below (see Equation 1). It can be easily implemented when the IV and the outcome are binary (Wang et al., 2017b).

$$P(Y_0, D_0|Z_0) + P(Y_1, D_0|Z_1) \leq 1,$$

$$P(Y_0, D_1|Z_0) + P(Y_1, D_1|Z_1) \leq 1,$$

$$P(Y_1, D_0|Z_0) + P(Y_0, D_0|Z_1) \leq 1,$$

$$P(Y_1, D_1|Z_0) + P(Y_0, D_1|Z_1) \leq 1,$$

Equation 1. IV inequality. Y_0 and Y_1 represents the outcome equals 0 and 1; D_0 and D_1 represents the treatment equals 0 and 1; Z_0 and Z_1represent the binary IV equals 0 and 1.

If this inequality does not hold, ER assumption and independence assumption are likely to be violated (Balke and Pearl, 1997, Wang et al., 2017b). Alternatively, one can find a subgroup where the IV does not affect the treatment. Since the effect from exposure has been excluded, association between IV and outcome in such subgroups implies the violation of exchangeability and ER assumption (Kang et al., 2013, Lipsitch et al., 2010). However, this method needs the subject knowledge to identify the possible subgroups (Labrecque and Swanson, 2018). Another way is to find a concomitant treatment of the treatment being investigated. If the IV is associated with the concomitant treatment, and at the same time the concomitant treatment affects the outcome, then the ER assumption is likely to be violated (Baiocchi et al., 2014).

2.2.3 Validation of independence assumption

An indirect way to validate the independence assumption is to check the imbalance of covariates across the levels of the IV. Imbalance in the observed covariates implies a possible association between the IV and unmeasured covariates. A commonly used approach is to calculate and describe the mean difference of covariates between levels of the IV (Davies et al., 2013c, Davies et al., 2020).

Together with the validation of ER assumption, another way is to validate the independence assumption is to find a sub cohort that has similar confounding structure to the study cohort. However, this subgroup is not exposed to the treatment under study, so it excludes the possibility that the IV affects the outcome via the exposure. If the IV has a direct effect on the outcome in this sub cohort, then the ER assumption and the independence assumption are unlikely to hold (Davies et al., 2017, Lipsitch et al., 2010).

2.2.4 Validation of treatment homogeneity and monotonicity assumption

The previous section explained two approaches to obtain parameter estimates of the treatment effect. One way is to assume treatment homogeneity among compliers, defiers, never takers and always takers. Another is to ensure there is no defier group (monotonicity assumption). In terms of falsification of the treatment homogeneity assumption, one potential approach is to examine the instrument strength across the level of the measured covariates (Rassen et al., 2009b). Moreover, a similar approach can also be used in the validation of the independence assumption (Labrecque and Swanson, 2018). Note that the treatment homogeneity assumption cannot be satisfied on both the additive scale and multiplicative scale (Hernán and Robins, 2006). Also, in some cases, treatment heterogeneity in epidemiology is considered likely (Poole et al., 2015). Alternatively, many researchers steer towards the validation of monotonicity assumptions (more details in section 2.2.5).

For binary IVs and binary exposures, no violation of monotonicity can be expressed as no defiers. However, the compliers groups, never takers and always takers group are impossible to be observed in the observational data separately (Swanson et al., 2015a). The monotonicity inequality proposed by Balke and Pearl can be used as an indirect empirical approach to validate monotonicity assumption when the exposures and outcomes are binary (Balke and Pearl, 1997) (see Equation 2). This equation is obtained based on the decomposition of conditional probability P(Y,D|Z)= P(Y|D,U)P(D|Z,U)P(U) and monotonicity assumption P(D=1|Z=1,U=u)≥ P(D=1|Z=0,U=u), proposed by Angrist and colleagues (Angrist et al., 1996).

$$P(\boldsymbol{Y}, D_1|Z_1) \geq P(\boldsymbol{Y}, D_1|Z_0)$$

$$P(\boldsymbol{Y}, D_0|Z_0) \geq P(\boldsymbol{Y}, D_0|Z_1)$$

Equation 2. Monotonicity inequality

2.2.5 Specific PPP IV considerations

As the only empirically verifiable assumption, the relevance assumption holds when the PPP IV is demonstrated to have a strong association with the treatment assignment. However, PPP is a proxy that may not directly reflect the association between the treatment and 'true' preference from the physicians. Weak associations between exposure and the IV leads to extremely biased results for PPP IV (Franklin et al., 2015). Independence assumption is likely to be violated in the case where patients with specific conditions may visit certain prescribers based on their prescribing preference; also referred to as 'doctor shopping'(Rassen et al., 2009a).Therefore, walk-in clinics and emergency room are more ideal for choosing PPP as IV (Potter et al., 2020). However, this is likely to be context-specific as not all medical treatment delivery outside walk-in clinics and emergency rooms is one homogenous grouping. For PPP IV, ER assumption is likely to be violated when physicians who prefer one particular type of treatment are more skilled in terms of delivering such treatment than physicians who prefer other type of treatment as they are more experienced (Baiocchi et al., 2014).

For PPP IV, monotonicity assumption is likely to be violated due to the reason that the PPP IV is defined by multidimensional elements (Swanson and Hernán, 2014). As mentioned before, the main obstacle of validating monotonicity assumption lies in the definition of a complier group and the defier group. For PPP IV, the definition of compliers from Table 3 is insufficient. The reason being one patient can be either a complier or defier under the treatment from a different physician. Since the treatment received by the same patients under a different physician cannot be observed in real life (counterfactual outcome), Swanson and colleagues (Swanson et al., 2015a) conducted a survey with physicians with a hypothesised group of patients to observe the counterfactual treatment received by the patients (Swanson et al., 2015b). They also undertook a pilot study which included 53 physicians and 20 hypothetical patients. 17 of 20 patients received the treatment that is the opposite of the prescribing preference of the physician which indicates a violation of monotonicity assumption. For preference based IV, surveys are an essential tool to validate the monotonicity assumption and are straightforward to conduct by asking the treatment provider about their decision (Swanson et al., 2015b). Since the deterministic monotonicity assumption may not be plausible for PPP IVs, Small and colleagues (Small et al., 2017) proposed a stochastic assumption which refers to the monotonicity within stratum formed by combinations of levels of covariates. Under the stochastic assumption, the LATE is a weighted sum of average treatment effects (Small et al., 2017). Swanson and Hernán (Swanson and Hernán, 2018) proposed that one needs to mention the estimated proportion of compliers in reporting IV estimates. However, for instruments such as PPP IV, the precise estimated proportion is unlikely to be observed. For PPP IV, one needs to establish an association between the unmeasured IV (for example, the true prescribing preference) and the surrogates for the unmeasured IV. If there is no subject-matter knowledge that can be used for building such an association, the proportion of the complier group will be bound between the denominator of the IV ratio ((E[Y|Z=1]-E[Y|Z=0])/(E[D|Z=1]-E[D|Z=0])) and 100% in the case of a positive association (Swanson and Hernán, 2018).

2.3 Type of treatment effects that IV method estimates

Before introducing the concept of the treatment effect, four core types of treatment effect need to be outlined. Rubin's causal model reveals that estimating the individual treatment effect is not possible as the counterfactual outcome cannot be observed within the same individual (Rubin, 1974). This led to the introduction of the concepts of average treatment effect (ATE), average treatment effect of treated group (ATT) and average treatment effect of untreated group (ATU). IV method yields the estimate of the local average treatment effect (LATE) (Fang et al., 2010) which is the average treatment effect among the patients whose treatment assignment is totally determined by the IV.

The potential differences between ATE, ATT and LATE are mainly related to treatment effect heterogeneity indicating patients with certain characteristics more likely differ in the amount they get benefit from the same treatment. However, if the treatment effect heterogeneity does not exist between the treatment group and control group, the estimation of ATE, ATT, ATU and LATE will be the same (Brooks and Fang, 2009). Normally, patients who tend to get health benefits are more likely to be treated, then ATT > ATU. The estimation from LATE falls between ATT and ATU as it estimates the treatment effect of a blend of the treated and untreated groups (Angrist, 2004). LATE can only estimate the treatment effect of compliers.

2.4 Estimation methods for IV method

In Table 4, I summarise the characteristics of commonly used estimation methods in IV method. The simplest estimation method for the IV estimates is the Ratio estimator (RE), also known as Wald estimator (Wald, 1940). This approach is suitable for single, binary IV. Unlikely RE, the two-stage least square (2SLS) method can adjust for covariates (i.e., potential measured confounding variables). 2SLS provides a consistent estimate of LATE when IV assumptions are hold. In the case of binary outcome, IV and exposure, the linear probability model (LPM) is equivalent to RE and 2SLS and provides estimated treatment effect on the risk difference scale. Two-stage predictor substitution (2SPS) is a non-linear

extension of 2SLS where the first stage of 2SPS non-linear regression on exposure and IV. The second stage regression uses predicted results from the first stage regression as an additional covariate. However, 2SPS and 2SLS generate similar results in the case of numerical exposures and outcomes (Klungel et al., 2015). Two-stage residual inclusion (2SRI) is another two-stage method in which the first stage of 2SRI is the same as 2SPS. In the second stage, it uses the residual from the first stage instead of model predictions from the first stage as an additional covariate.

Two-stage logistic regression (2SLR) is a two-stage regression model that uses logistic regression in both stages. This model is unable to provide a causal OR estimate due to the noncollapsibility of OR. This is also the case for 2SRI and 2SPS when the IV estimates are ORs. For binary outcomes, another two-stage method, bivariate probit model (BVP), is sometimes favoured. It provides the probit coefficients which can be approximated to an OR. Structural mean model (SMM) is a semi-parametric model that uses g-estimation to estimate causal parameter based on the conditional mean independence (Robins, 1994). Generalised method of moments (GMM) is a non-parametric method and can be used in over-identified models (i.e., IVs outnumber the endogenous exposures). Since it is not constrained by parametric assumptions, GMM is generally more efficient than BVP and 2SLR when the outcomes and exposures are binary (Klungel et al., 2015). However, in my review paper that includes 18 CERs which used PPP IV from 2005 to 2020, 11 of 18 studies conducted 2SLS. 1 of 18 studies conducted 2SRI and 2SLR respectively (Zhang et al., 2022). 2SLS is the most used estimation method on this topic.

Types of IV estimation method	**Notation**
Ratio estimator (RE)/ Wald estimator	$\frac{E[Y\|Z=1]-E[Y\|Z=0]}{E[D\|Z=1]-E[D\|Z=0]}$
2SLS	Consistently estimate LATE in the case of single IV and linear models
2SLR	• Logistic regression in both stages • Subject to noncollapsibility
2SPS	• Non-linear extension of 2SLS • Subject to noncollapsibility
2SRI	• Non-linear extension of 2SLS • Proved to be more consistent than 2SPS • Subject to noncollapsibility
SMM	• Semi-parametric modelling • Using G-estimation
BVP	• Model the probability of receiving the treatment and the outcome • Estimate probit coefficients • Perform better than 2SLS for binary outcomes and treatments
GMM	• Moment based • More consistent than 2SLS • Suitable for binary outcome

Table 4. Summary of common IV estimation methods

2.4.1 Specific PPP IV considerations

In the case of implementation of PPP IV, the choice of estimation methods should depend on the treatment effect of interest. Under the treatment effect homogeneity assumption, the estimation methods mentioned above provide estimates of ATE. Under the monotonicity assumption, all methods except for SMM provide estimates of LATE. While SMM provides estimate of ATT with the no effect modification assumption (NEM) holds (Klungel et al., 2015, Clarke and Windmeijer, 2012). The NEM assumption indicates that the IV does not modify the effect of treatment on the outcome (E [Y(1)-Y(0)|X=1,Z=1]=E[Y(1)-Y(0)|X=1,Z=0]) (Hernán and Robins, 2006). Another nonparametric method is the marginal treatment effect (MTE) which is proposed by Heckman and Vytlacil (Heckman and Vytlacil, 2005). Local IV (LIV) parameter estimation (Heckman and Vytlacil, 1999), integrated with propensity scores, is a common way to bound treatment effects within a MTE model.

2.5 Discussion

Although many existing literature suggests cautionary implementations of preference-based IVs (Ionescu-Ittu et al., 2009, Garabedian et al., 2014, Franklin et al., 2015), and the need to be assessed on the case-by-case basis (Swanson and Hernán, 2018), I still believe in the utility of PPP IV for use in CER. The natural variation from PPP can be a close approximation to the ‘flipping coin’ assignment mechanism in randomised comparative effectiveness studies, which cannot be replicated by other forms of IV, such as distance and calendar time. A further strength is that preference-based IVs tend to be strongly associated with the treatment assignment. Also, the preference-based IVs are less likely to be associated with the outcomes (violating the exclusion restriction assumption) in many scenarios. I have pointed out that the relevance assumption is the only assumption that can be validated empirically, and in evidence across many applied studies, the PPP IV is shown to reduce covariate imbalance, therefore less likely to violate the independence assumption.

The pitfalls of using PPP IV mainly lie in the validation of the monotonicity assumption. As mentioned earlier, one needs to assume monotonicity assumption hold to achieve parameter estimates of LATE. However, the validation of monotonicity assumption is usually overlooked. When the questionnaire survey is not feasible, I recommend researchers validate the monotonicity assumption using indirect approaches, such as monotonicity inequality or reporting sensitivity analysis of the violation of monotonicity assumption and the estimated proportion of compliers/defiers. If the validation of monotonicity is difficult to achieve, while other assumptions are believed to hold, one can report the bound of the estimated treatment effect from nonparametric approaches, such as Balke and Pearl proposed (Balke and Pearl, 1997). Another noticeable limitation of IV methods is the weaker statistical power compared with the conventional methods, such as multivariable regression, propensity score. For that, stronger IV and larger sample size are preferred (Martens et al., 2006).

2.6 Conclusion

This chapter provides an introduction and critical commentary of CERs with emphasis on using PPP as an IV. I encourage investigators to access the assumption of IV for their specific applications when then use the PPP IV. use PPP as IVs when unmeasured confounding bias is likely. However, they should be cautious in terms of the complexities arising from the monotonicity assumption and strength of the association between the treatment and the IV.

Chapter 3. Physician's prescribing preference instrumental variables: comparative effectiveness research should consider methodological insights from simulation studies.

3.1 Authorship, and publication details

Zhang, L., Lewsey, J., & McAllister, D. A. (2022). Comparative effectiveness research considered methodological insights from simulation studies in physician's prescribing preference. Journal of Clinical Epidemiology, 148, 74-80.

3.2 Abstract

Objective:

To review CER using PPP as IV in pharmacoepidemiology, and to review methodological studies that use simulation to evaluate the performance PPP IV in CER.

Study design and setting

We conducted a review of CER using PPP IV as well as studies evaluating the use of PPP IV by using simulation methods. We searched Ovid, PubMed, and Google Scholar databases from 2005 to 2020.

Results

We identified 6 simulation studies and 18 CERs. The simulation studies explored the most suitable ways for using PPP IV in different settings (outcome types, sample size, the prevalence of outcomes) which can be useful guidance for using PPP IV in CER. The CERs identified show heterogeneity in terms of validation assumptions, estimation methods and sample size. Not all applied studies utilised the methodological insights from the simulation studies. However, they all concluded that PPP is a valid instrumental variable.

Conclusion

Future CER should consider a range of methodological issues to improve the validity of findings when using PPP IV. Specifically, studies should consider the impact of different choice of statistical methods, forms of proxy for measuring preference, time-varying exposures, and the type of outcome.

Key words: review, pharmacoepidemiology, instrumental variable, physician's prescribing preference, comparative effectiveness research, simulation studies

What is new?

- This article reviews applied and simulation studies that use physician's prescribing preference as instrumental variables in pharmacoepidemiology between 2005 and 2020.
- In this review, the applied studies that use physician's prescribing preference as an instrumental variable do not always report their results to a standard that systematic reviews and simulation studies suggest.
- Applied studies using physician's prescribing preference as an instrumental variable should consider methodological insights from simulation studies to inform study designs.

3.3 Introduction

In comparative effectiveness research (CER) using observational study designs, residual confounding is the one of the most important challenges. There are well-established methods that focus on reducing covariate imbalance, such as multivariable-adjusted regression and propensity score methods (Rosenbaum and Rubin, 1983).. However, these methods assume no unmeasured confounding after such adjustment. Originally from econometrics, the instrumental variable (IV) method can be used to address unmeasured confounding (Angrist and Imbens, 1995).

Physician's prescribing preference (PPP) is a commonly used instrument (Brookhart et al., 2007). It exploits naturally occurring variation which makes the treatment assignment in observational studies using heath datasets closer to that in randomised controlled trials. The PPP is a latent variable and relies on proxy/surrogate measurement. The most common proxy for PPP is the most recent prescription made by the same physician for patients with the same symptoms, also referred to as 'prior one' prescription (Brookhart et al., 2007). Alternatively, PPP can be defined as the proportion of one particular drug under study prescribed among all the previous patients of the physician (Ionescu-Ittu et al., 2009). Further, PPP can be constructed at higher levels of aggregation, such as general practice (a local grouping of doctors which is found in the UK and other similar health systems) or other regional levels (Garabedian et al., 2014).

In the past 15 years, there had been research papers which introduce IV methods in general and the use of PPP IV specifically (Hernán and Robins, 2006, Lousdal, 2018, Baiocchi et al., 2014). Chen and colleagues conduct a systematic review to synthesise drug research studies using IV to see whether it can be a valid approach for tackling unmeasured confounding bias (Chen and Briesacher, 2011). Davies and colleagues also conduct a systematic review of using IV methods and reporting of IV results (Davies et al., 2013b). However, these cover common types of IV used in epidemiological research and so did not cover some issues of particular importance to PPPs. This paper reviewed the scientific literature of CERs using PPP IV to compare how methods are employed in applied studies to those studies introducing important methodological considerations.

3.4 Method

3.4.1 Inclusion and Exclusion Criteria

To identify simulation studies, the inclusion criteria were: 1) Use PPP as IV in a hypothesised drug comparison study; 2) Published between 2005 to 2020.

To identify observational CERs using PPP IV, we conducted a literature search using Ovid, PubMed, and Google Scholar with the following inclusion criteria: 1) Use prescription drugs as exposure; 2) Compare the effectiveness of two drugs; 3) Use PPP as IV; 4) Published between 2005 to 2020. The exclusion criteria were: 1) Review paper; 2) Clinical trial; 3) Abstract or book.

3.4.2 Search terms

The search terms used in this review were: instrumental variable AND prescribing preference OR medication OR treatment for both applied and simulation studies. We distinguished between these two types of study according to whether they used simulated data or routinely collected data.

3.5 Results

Using the search terms, we found 1192 records in Ovid, PubMed and Google Scholar (see Figure 2). We first excluded studies that were not comparing treatment effectiveness. Then we excluded studies that are irrelevant to IV methods. Finally, we deleted those studies that used the IV method but not specifically PPP IV. The remaining 18 applied and 6 simulation studies are included in the subsequent review.

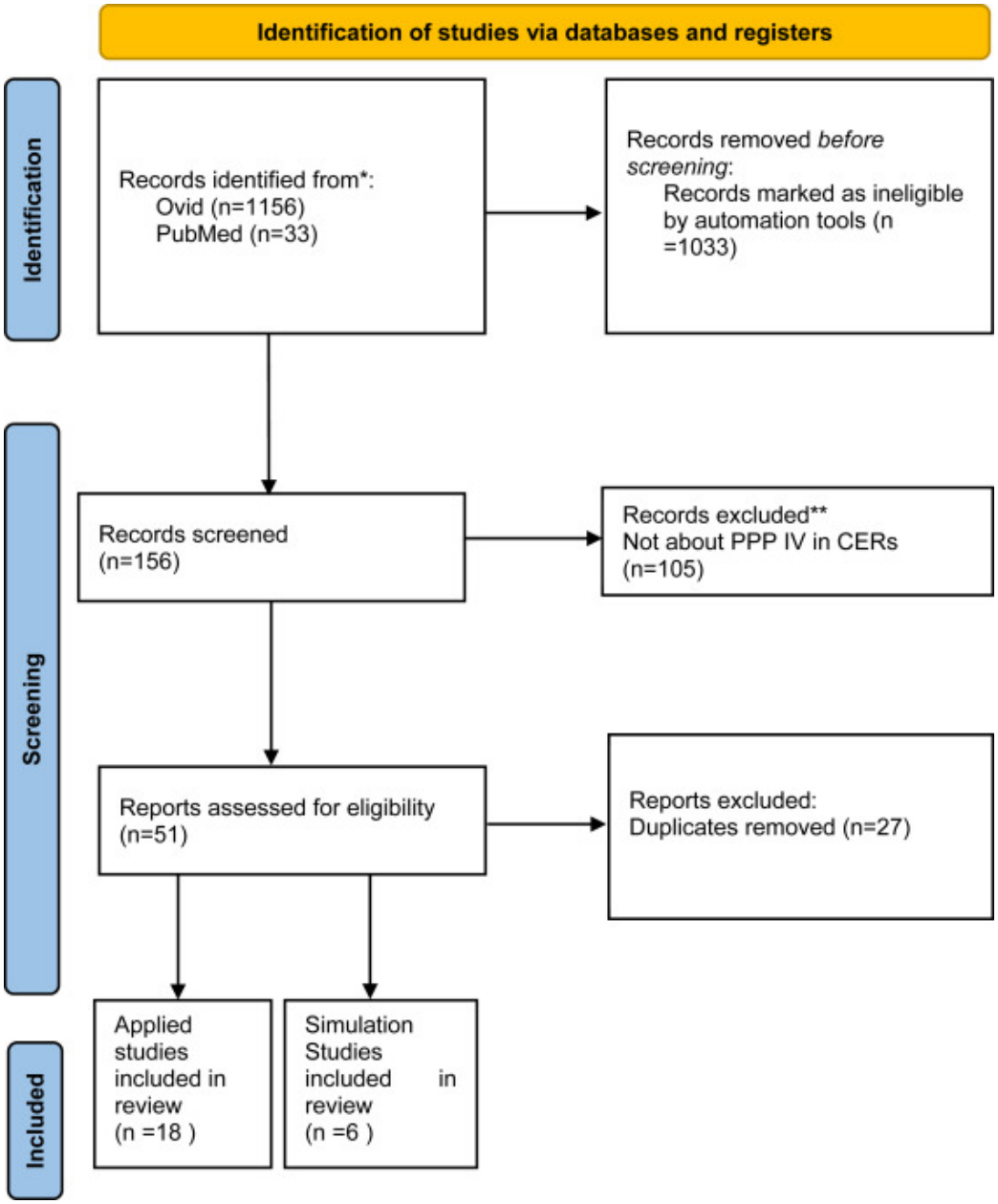


Figure 2. Flow chart of literature search

We identified 6 simulation studies and 18 applied studies (see Table S1 and Table S2 in supplementary material for Chapter 3).

3.5.1 Summary of the simulation studies

All identified simulation studies aimed to get better understanding of using PPP IV as a method of reducing unmeasured confounding bias in pharmacoepidemiology (see Table S1 in supplementary material). For that, they formed their data to simulate hypothetical studies comparing the treatment effectiveness or the risk of adverse event in the context of large datasets. The PPP was built on the basis that all three assumptions are met. In general, better performance means lower variance (smaller standard deviation) and lower bias. Simulated data facilitates the comparison between the IV estimate with the 'true' estimate and use RMSE, relative bias, and coverage rate to measure the bias quantitatively. Most of these studies mention the strength of the association between exposure and the instrument. They emphasise the association between instruments and exposures needs to be strong enough to implement unbiased instrumental variable analysis. Ionescu-Ittu and colleagues used the proportion of exchangeable group to represent the strength of this association (Ionescu-Ittu et al., 2009). A further study (Uddin et al., 2014) also focuses on the strength of PPP IV but in a more specific way by defining boundaries for weak instruments (e.g., Pearson's correlation coefficient < 0.15 or odds ratio < 2). Also, the limitations of using IV methods, including the weak instrument and limited sample size can be shown in a more specific way (Uddin et al., 2016a).

3.5.2 Summary of the applied studies

3.5.2.1 The Construction of Proxy for Physician's Prescribing Preference

In terms of constructing PPP IV, there are two major types of variables in the applied studies: binary (the most recent prescription made by the same physician – 'prior one') and numerical (the proportion of patients who were prescribed drug of interest). The variance of binary outcomes are more likely to be inflated than continuous outcomes under the same settings (Ionescu-Ittu et al., 2009). Unlike the other studies, Koladjio and colleagues compared the results of GMM and 2SRI which adds new knowledge on IV analysis using non-linear regression models (Koladjo et al., 2018). Most of these studies used the most recent prescription, or the prior one prescription of the same physician (Davies et al., 2013a, Taylor et al., 2017, Davies et al., 2013c, Schneeweiss et al., 2007, Schneeweiss et al., 2006, Chen et al., 2014). However, no study provided a rationale for the choice of form of PPP.

Some papers conducted sensitivity analysis by comparing the estimates from prior one and prior *n* prescriptions, such as prior 7 and prior 20 prescriptions (Davies et al., 2013a, Davies et al., 2018, Davies et al., 2020). The other form of PPP IV is the proportion of one drug prescribed by the physician among all the previous patients which makes the IV a numerical variable (Boef et al., 2016b, Kollhorst et al., 2016). The magnitude of a preference can also be defined using the proportion of one drug prescribed by the physician dichotomized at the median (Kuo et al., 2012, Secemsky et al., 2017). In higher levels of aggregation, studies do not account for the prescribing date but include the proportion of one treatment among all prescriptions (Uddin et al., 2016b).

3.5.2.2 Estimation method in different settings

Although 2SLS is widely used in pharmacoepidemiology to account for unmeasured confounding for different types of outcome variable (Hernán and Robins, 2006). Literature shows that using the 2SLS estimator for binary outcomes may lead to a biased estimate (Ertefaie et al., 2017). Table 5 is a summary of the most suitable estimation methods in different settings. In the review, although all CERs have binary or time-to-event outcomes, only 7 of 18 (39%) use non-linear models instead of 2SLS. Regarding time-to-event outcomes, two-stage regression models that include the instrumental variable are proposed (Tchetgen et al., 2015, Martínez-Camblor et al., 2019). The first stage is linear regression on the exposure and instrumental variable with measured covariates adjusted for. The second stage is a Cox proportional regression model that adjusts the survival probability from the first stage (Boef et al., 2016b).

Estimation	**Suitable scenarios**
Two-stage Linear Regression	Linear model (Hernán and Robins, 2006)
Generalized method of moment (GMM) IV	Binary outcome (Klungel et al., 2015)
Two-stage residual inclusion(2SRI)	Binary outcome (Terza et al., 2008b) Numerical outcome (Zhang et al., 2018) Time-to-event outcome (Terza et al., 2008a)
Two-stage predictor substitution (2SPS)	Binary outcome (Terza et al., 2008a), time-to-event outcome (Cai et al., 2011)
Local Average Treatment Effects (LARF)	Binary treatment and binary instrument (Zhang et al., 2018)
IV Cox Regression	Time-to-event outcomes (Boef et al., 2016b)
GLM adaptions on IV (e.g. IV Probit / IV logistic Regression)	Binary outcome (Rassen et al., 2009b)

Table 5. Summary of estimation methods in different settings

3.5.2.3 Validation of assumptions

In terms of validating IV assumptions (see Table 6), most of the applied studies assessed the strength of the association between IV and the exposure (the relevance assumption). One of the most common ways is to calculate partial F-statistics on the first stage of regression (Kuo et al., 2012, Davies et al., 2013c, Davies et al., 2013a, Nelson et al., 2013, VanDyke et al., 2013, Chen et al., 2014, Boef et al., 2016b, Kollhorst et al., 2016, Davies et al., 2020). The rule of thumb for a strong enough instrument is if the F-statistics is greater than 10 (Stock and Yogo, 2002). Comparing the percentage of the actual treatment that match the prescribing preference is a more intuitive way to see if IV predicts the actual treatment (Brookhart et al., 2006, Secemsky et al., 2017). There are also studies that fit regression models on exposure and IV to examine whether the association is statistically significant (Chen et al., 2014, Schneeweiss et al., 2007). Of these different approaches, the F-statistics has the advantage that it accounts for the strength of association as well as being sensitive to sample size. Note, however, that a large F-statistics only indicates that weak instrument bias is unlikely, it does not follow that there will be enough statistical power to adequately test treatment effectiveness.

The exclusion restriction and independence assumption cannot be tested empirically (Hernán and Robins, 2006). However, many studies compared the covariate imbalance by actual treatment and by PPP with an intuition that if PPP is less associated with measured confounders then it will also be less associated with unmeasured confounders (Brookhart and Schneeweiss, 2007). Likewise, there are studies that reported the reduction of covariate imbalance using Mahala Nobis distance (Davies et al., 2013a), bias component plot (Davies et al., 2018) and Prevalence difference ratio (PDR) (Davies et al., 2013a, Davies et al., 2013c). The exclusion restriction assumption which indicates that the IV does not affect the outcome directly has been overlooked in the identified studies of the review. Most studies did not mention or simply explained it using intuition that the preference of prescribing is not likely to influence the outcome (Kuo et al., 2012).Of the studies identified, only one (Kollhorst et al., 2016) explored further and constructed an adjusted logistic regression model on IV and the outcome to examine whether PPP can affect the outcome.

Study	Validation of assumptions
(Brookhart et al., 2006)	a. The probability of receiving COX-2 when the prior one prescription is COX-2
(Schneeweiss et al., 2006)	a. Compares the percentage of the same physician prescribed the same drug as the prior one and the that of different drug. b. The reduction of imbalance in covariates.
(Schneeweiss, 2007)	a. Association between the instrument and the exposure: OR: 6.1, 95% CI (5.8–6.4).
(Schneeweiss et al., 2008)	N/A
(Davies et al., 2013a)	a. Cluster-robust F-statistics b. The risk difference of confounders on the level of actual treatment and on the level of IV. And the reduction of Mahala nobis distance and the prevalence difference ratio
(Kuo et al., 2012)	a. Partial F statistics
(Davies et al., 2013c)	a. Partial F-statistics and R-square of linear regression on exposure and instrument b. Prevalence difference ratio
(Davies et al., 2018)	a. Partial F-statistic b. Bias component plot
(Davies et al., 2020)	a. Partial F-statistics
(VanDyke et al., 2013)	a. Partial F-statistic and sensitivity analysis.
(Nelson et al., 2013)	a. Partial F-statistic
(Taylor et al., 2017)	N/A
(Kollhorst et al., 2016)	a. Partial F-statistics and R-square, the square of the partial Spearman correlation coefficient, b. Partial F-statistics for the regression model on instrument and three forms of IV. c. Adjusted logistic regression on IV and outcome
(Boef et al., 2016b)	a. Partial F-statistics and R-squared
(Chen et al., 2014)	a. Partial F-statistics and the first stage linear probability
(Secemsky et al., 2017)	a. The percentage of actual treatment with a high preference for Bivalirudin.
(Uddin et al., 2016b)	a. Point bi-serial correlation (r) for binary exposure and continuous IV; ORs for binary exposure and IV. c. Standardized difference and multivariate Mahala nobis distance assess the imbalance of covariates
(Suh et al., 2012)	a. F-statistics of the first stage of regression b. Reduction of imbalance in covariates.
(Walker et al., 2020)	a. F-statistics of first stage of regression and the mean association between exposure and instrument b, c. Bonet's instrumental variable inequality tests.[1]

Table 6. Validation of assumptions.

[1] *a. The validation of relevance assumption of IV. b: The validation of the exchangeability assumption of IV. c: The validation of the exclusion restriction assumption of IV.*

3.6 Discussion

In this paper we have reviewed both the applied and methodological literature that use PPP IVs to address unmeasured confounding in CER. We found that the methodological insights from the simulation studies were not always being considering, or at least not reported on, in the applied studies.

Although some studies argued that 2SLS can be used as is asymptotically unbiased (Ionescu-Ittu et al., 2009), researchers should endeavour to compare the results from 2SLS and that from other estimation methods to increase the robustness of the study. In terms of different forms of proxy for PPP, simulation studies have compared the performance of numerical and binary PPP formulations. Although some researchers concluded that the proportion form of PPP serves as a good proxy for PPP (Koladjo et al., 2018), others held a view that instantaneous preference is better which is consistent with the earlier definition of PPP IV (Brookhart and Schneeweiss, 2007). Further, longer prescription histories used in the PPP formulation leads to dropping those prescribers with few prescription records and dropping long 'look-back' periods from the analysis cohort, so a trade-off must be made. Given that the performance of PPP IV can depend on multiple elements, such as the rarity of outcome (Ionescu-Ittu et al., 2009), sample size (Boef et al., 2014) as well as the estimation methods, a proper proxy for PPP needs more exploration than tends to occur in the applied literature to date. We suggest that researchers should present their IV estimates using more than one form of proxy as sensitivity analyses for possible violation of IV assumptions. Unlike in the simulation studies, the identified applied studies in the review did not account for or consider time-varying preferences. Abrahamowicz and colleagues (Abrahamowicz et al., 2011) assumed the preference would switch (from preferring drug A to drug B) at a certain point in time. They include the time factor by using the prescription history as the proxy for PPP.

It should be noted that simulation studies cannot provide insight for validation of all IV assumptions. Although the exchangeability and exclusion restriction assumptions cannot be empirically verified, researchers should consider them in study reporting. Some studies have proposed a falsification strategy which may be a possible solution for assumption checking (Labrecque and Swanson, 2018). Further research may focus on explicitly how such a strategy can be used to help validate independence assumption and exclusion restriction assumptions.

Of course, simulation studies can only 'mimic' real life data sets. The simulation studies we reviewed build hypothetical studies to compare the effectiveness of two drugs. In some instances, the models' parameters are set and varied without providing justification which is not consistent with the best practice for simulation studies (Morris et al., 2019) which is partly due to these simulation studies were conducted before such guidance. Further, the results of simulation studies often favour the new method that is being proposed/introduced which may be overstating what happens in real life data sets. Although there are no specific standards, some review papers have made suggestions on the reporting of IV studies. Brookhart (Brookhart et al., 2007) and Davies (Davies et al., 2013b) proposed guidelines for reporting IV analyses results. More specifically, Swanson and Hernán (Swanson and Hernán, 2013) proposed a flowchart for reporting IV analyses of single binary non time-varying IVs including a classification between average treatment effect (ATE) and local average treatment effect (LATE). Baiocchi and colleagues (Baiocchi et al., 2014) highlighted the exploration of concomitant treatment and sensitivity analyses. Jackson emphasised properly demonstrating the confounding bias by bias component plots (Jackson and Swanson, 2015).

Additional related issues for PPP IV are those of weighting and matching. The instrumental propensity score (IPS) is used which is the conditional probability of an IV given pre-treatment covariates as weights in the regression models (Tan, 2006). Cheng and colleagues further explored the IPS approach by implementing it in subclassifications and semi-parametric models to gain treatment effects of subgroups (Cheng and Lin, 2018). Matching methods have been used to increase the strength of weak IVs. Baiocchi and colleagues (Baiocchi et al., 2010, Baiocchi et al., 2012) proposed a near-far approach to create pairs with similar covariate distributions and large differences in IV value to mimic randomized trials. IPS can also be used in a full matching approach (Kang et al., 2013). Although not central to the design of PPP IV studies, we recommend that using these methods should be considered in CER, especially when the proportion of the complier group is relatively low.

We evaluated the 18 applied CER papers according to whether they account for insights from simulation studies by four criteria: 1) whether they compare different forms of PPP IV; 2) whether they compare IV methods with conventional method; 3) whether they compare different estimation methods; 4) whether they consider time-varying preferences. In summary, 5 of 18 studies (28%) compared different forms of PPP IV, and 18 of 18 studies compared the IV methods with conventional methods. However, none of the studies considered time-varying preferences or compared different estimation methods which are often covered by the simulation studies.

3.7 Conclusion

In terms of exploring the suitability of the PPP IV method, simulation studies are flexible, easy to use and have potential to make recommendations for good practice in the applied setting. We have shown that currently not all applied studies use simulation study results to guide their use of the PPP IV method. Applied studies using PPP as an IV should consider methodological insights from simulation studies to inform study designs. A brief checklist (see Table 7) is presented for researchers who are interested in applying PPP IV.

1. Preregister IV definitions
2. Report the strength of IV
3. Check the balance of confounders based on IV
4. Compare different formulations of PPP instrument.
5. Compare the IV method results with those from conventional methods, such as multivariable adjusted regression or propensity score approaches.
6. Use different estimation methods which have different underlying assumptions to explore the robustness of the study findings.
7. Consider accounting for time-varying prescribing preference (or provide a clear rationale for assuming that the physician's prescribing preference is time-fixed across the period of study).

Table 7. Checklist for the researchers who are interested in using PPP in CERs

Chapter 4. Using Physician's prescribing preference as an instrumental variable to compare the effectiveness of diazepam and chlordiazepoxide hydrochloride for patients diagnosed with alcohol intoxication and harmful use.

4.1 Publication details

This article had not been submitted for publication.

2.3.1 Specific for PPP IV considerations

For PPP IV, the identification of treatment effect strongly relies on the three core concepts: 1) definition of PPP; 2) validation of the monotonicity assumption; 3) strength of the IV. For PPP IV in particular, the monotonicity assumption is not plausible in many settings (Swanson and Hernán, 2014). Further, even with a strong preference, physicians cannot ensure that there are strictly no defiers. In addition, different physicians at the same level of prescribing preference will not treat patients in the exact same ways suggesting that the monotonicity assumption is likely to be violated (Swanson and Hernán, 2014). Although different definitions of PPP IV imply treatment effects from different subsets of compliers, stronger IVs are always preferred since they can reduce the variance of IV estimates (Ionescu-Ittu et al., 2012). Sensitivity analyses of monotonicity assumptions suggest stronger IVs are more robust to the violation of monotonicity (Baiocchi et al., 2014).The validation of the monotonicity assumption will be discussed further in the following section.

4.2 Data used in this chapter.

Data used in this chapter was applied for under the application number 1718-0238 to Public Health Scotland's national safe haven (eDRIS). The eDRIS user agreement is attached in the appendix. To form the outcome as the rehospitalisation and death, Scottish Morbidity record (SMR01) was also linked.

Data tables and variables that are included in the regressions are listed in the table below. The description of variables can be found at https://www.isdscotland.org/Health-Topics/Prescribing-and-Medicines/Prescribing-Datamarts/docs/PIS_fields_for_researchers_v5_eDRIS%20Guidance.pdf

Table	Variable	Brief explanation of some variables
PIS_2011_.dta PIS_2012_.dta PIS_2013_.dta PIS_2014_.dta PIS_2015_.dta PIS_2016_.dta PIS_2017_.dta PIS_2018_.dta PIS_2019_.dta	patientid patgendercode Age_paid_date prescriberprofessionalnoanon prescdate pibnfrootdrugdescription	prescriberprofessionalnoanon : Prescriber professional number, unique for each physician.
SMR01	patiendid main_condition other_condition_1 other_condition_2 other_condition_3 other_condition_4 other_condition_5 admission_date discharge_date_	

4.3 Abstract

Background

Alcohol use disorder (AUD) has been one of the major health concerns in Scotland. Multiple pharmacological treatments are available for treating AUD, for example benzodiazepines for alcohol withdrawal syndrome (AWS). However, there is limited evidence of the effectiveness of pharmacological treatments for alcohol intoxication and harmful use (AIH) which are outcomes that often precede AWS. In this chapter, I investigated the effectiveness of diazepam and chlordiazepoxide hydrochloride in preventing AIH rehospitalisation and AIH death from a real-world perspective using an observational study design. This chapter is a CER study using unselected data for the whole of Scotland.

Method

Data used in this chapter came from Scottish Prescription Information System (PIS) dating back from 2010 to 2019 linked to hospitalisation and death records, with an index cohort defined by first prescription of diazepam or chlordiazepoxide hydrochloride following a AIH hospitalisation (within 1 year time window). The outcome measures under study are AIH rehospitalisation and AIH death. Statistical methods consist of conventional statistical methods, including OLS and Cox proportional hazards regression, as well as the IV approaches, including 2SLS and IV Cox regression and to addressing potential unmeasured confounding issue. PPP IV was constructed using the prescription history of the physicians.

Results

The estimated hazard ratio comparing diazepam with chlordiazepoxide hydrochloride for AIH rehospitalisation from multivariable Cox proportional hazards regression is 0.772 (95% CI: 0.656-0.907) and the estimated hazards ratio of AIH death is 0.695 (95% CI: 0.51-0.946). The estimated hazard ratio from IV approach is between 0.538 and 0.826 with variation due to the different PPP formulations, with 95% CIs crossing the null value in each case. PPP IVs constructed with a longer prescription history were more likely to be stronger IVs.

Conclusion

This nationwide CER from Scotland provides evidence to support diazepam performs better in reducing the risk of AIH rehospitalisation and AIH death in comparison to chlordiazepoxide hydrochloride. This evidence comes from methods that do and do not try to address unmeasured confounding. However, the sample size leads to large imprecision in the PPP IV estimates.

Keywords: alcohol intoxication harmful use, physician's prescribing preference, instrumental variable, diazepam, chlordiazepoxide hydrochloride

4.4 Introduction

4.4.1 Background

Alcohol use disorder (AUD) is a major health concern worldwide (World Health Organisation, 2014). About 13.5% of deaths and disability among people aged 20-39 years are due to harmful use of alcohol (WHO, 2022). In Scotland, there were 33,015 AUD-related general acute hospital stays and 2,109 admissions in psychiatric hospitals reported from 2020 to 2021 (Scotland, 2021). Further, there were 1,190 alcohol-related deaths in Scotland in 2020, the highest rate in the UK at 21.5 per 100,000 people (Office for National Statistics, , 2021).

4.4.2 Pharmacological treatment of AUD (alcohol withdrawal syndrome)

Benzodiazepine has shown to have protective effects against AUD syndrome, especially for a diagnosis of alcohol withdrawal syndrome (AWS) (Amato et al., 2010, Mayo-Smith, 1997, Bahji et al., 2022). In a systematic review of benzodiazepines, chlordiazepoxide hydrochloride generally had better results but without achieving statistical significance (Amato et al., 2010). Among the four most commonly prescribed benzodiazepines (lorazepam, chlordiazepoxide hydrochloride, oxazepam, and diazepam), diazepam takes the shortest time to peak in its effect (Weintraub, 2017, Lee et al., 2019). After a comprehensive review of literature, Weintraub concluded that diazepam should be preferred as treatment for patients with mild to severe alcohol withdrawal in most cases (Weintraub, 2017).

4.4.3 Pharmacological treatment of AUD (with history of alcohol intoxication and harmful use)

Alcohol intoxication, which refers to a large ingestion of alcohol (Vonghia et al., 2008), as along with alcohol harmful use, which refers the a drinking pattern that leads to alcohol-related health problems, are two major syndromes of AUD. However, there is a lack of RCT evidence about which benzodiazepine treatments are better in reducing risk of alcohol intoxication and harmful use (AIH) events.

Normally, the treatment for AIH is more likely to be brief behavioural intervention rather than pharmacological treatment. This usually comprises a short counselling session aiming to change addictive behaviours (Barnes and Samet, 1997). Such interventions have proven to be effective in reducing alcohol consumption in multiple RCTs and observational studies (Reid et al., 1999, Harris et al., 2014, Walton et al., 2010). However, when pharmacological treatment is provided, diazepam and chlordiazepoxide hydrochloride are often prescribed for patients with AIH hospitalisation history. Diazepam and chlordiazepoxide hydrochloride are the commonly used benzodiazepines to control AWS (Jauhar and Anderson, 2000). Diazepam is more likely to be prescribed to patients with AWS seizures and with quicker take effect (Weintraub, 2017, Schmidt et al., 2016). While chlordiazepoxide hydrochloride has lower abuse potential (Jauhar and Anderson, 2000). Despite the fact these medications are recommended by NICE guidelines as pharmacotherapies for treating alcohol withdrawal (National Collaborating Centre for Mental Health, , 2011), there are not many studies except for a small scale pilot study that compared the effectiveness of diazepam and chlordiazepoxide in treating AWS (Jauhar and Anderson, 2000), with their results favouring daily single dosage of diazepam as the treatment for alcohol detoxification. Considering a large proportion of patients with severe AIH may develop AWS when they stop drinking (Day and Daly, 2022), research is needed to compare the effectiveness of pharmacotherapies after a severe AIH event, for example, one that was severe enough to result in hospitalisation.

CERs that use routinely-collected health data sets offers the opportunity to provide real-world evidence on the effectiveness of pharmacotherapies, however, in such observational studies, unmeasured confounding (confounding by indication) is a severe threat to the internal validity of findings. Increasingly used in practice, IV method has been shown to be a useful approach to directly address unmeasured confounding concerns. Among all the potential types of IV, PPP has been shown to be a valid instrument in many settings (Brookhart et al., 2007).

According to real-world data: Scottish prescribing information system (PIS) (Alvarez-Madrazo et al., 2016), the two most commonly prescribed medications for patients in Scotland with AIH hospitalisation records are diazepam and chlordiazepoxide hydrochloride from 2010 to 2019. The aim of this study is to compare the effectiveness of diazepam and chlordiazepoxide hydrochloride in terms of preventing AIH related events. Specifically, I used PPP as IVs to compare the effectiveness of diazepam and chlordiazepoxide hydrochloride in preventing AIH rehospitalisation and AIH death for patients prescribed diazepam or chlordiazepoxide hydrochloride after their first-time AIH hospitalisation. The IV results are compared to the results from multivariable regression, a method that can only account for measured covariates.

4.5 Method

4.5.1 Physician's prescribing preference as an instrumental variable

PPP is a latent variable which cannot be measured directly, but statistics derived from prescription history can be used as a proxy (Davies et al., 2013b). The calculation of PPP is based on physician and may vary by different physicians. The PPP IV used in this study is estimated as the proportion of diazepam prescribed by one particular physician in the preceding year after a patient has a first AIH hospitalisation. There are two forms of PPP: prior *n* and proportional (see Equation 3 and Equation 4).

$$Prior\ n\ IV: \frac{Total\ number\ of\ \ diazepam\ \ among\ prior\ n\ prescriptions}{n}$$

Equation 3. The calculation of prior n IV

$$Proportional\ IV: \frac{Total\ number\ of\ diazepam\ dates\ back\ 1\ year}{Total\ number\ of\ diazepam\ and\ chlordiazepoxide\ \ prescriptions\ dates\ back\ 1\ year}$$

Equation 4. The calculation of proportional IV.

The process of calculating the PPP IV is 1) identify the prescriber that issued the 'First AIH prescription' for the patients; 2) then, for that physician's ID, calculate the PPP IV. See Figure 3 for the details.

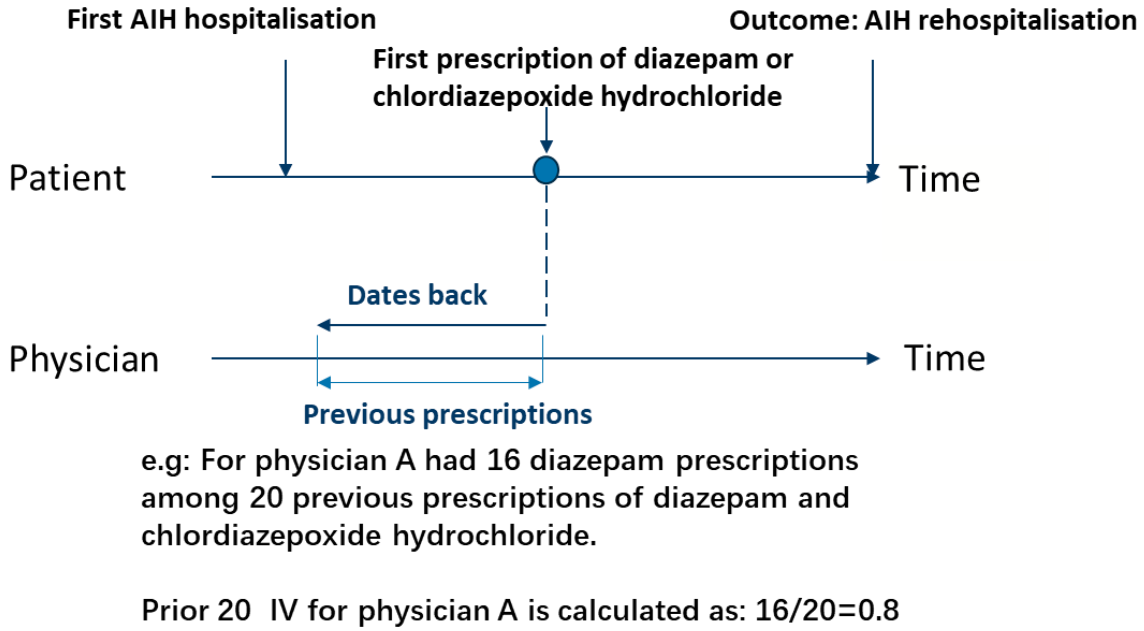


Figure 3. Construction of PPP IV

4.5.2 Validation of assumptions

I validated the relevance assumption using F-statistics of the first stage regression calculated in the diagnostic function inside the 'ivreg' function in 'AER' R package (Kleiber and Zeileis, 2008). The rule of thumb for identifying a weak instrument is F-statistics less than 10 (Stock and Yogo, 2002). Since the F-statistic is highly related to the sample size (see Equation 5), and sometimes misleading (Martens et al., 2006), I regressed the treatment on the PPP IV in logistic regression and used 'area under curve' (AUC) to measure the strength of the association between the treatment and the PPP IV.

$$F\ statistics = \frac{\rho_{ZX}^2(n-2)}{1-\rho_{ZX}^2}$$

Equation 5. F statistics; σ_X^2: The variance of the treatment; $\rho_{Z,X}^2$: The correlation between the treatment(X) and the instrumental variable (Z).

In order to make the validation of assumptions easier, many studies suggest dichotomising numerical IV (Uddin et al., 2016b, Secemsky et al., 2017). In this chapter, the numerical PPP IVs (which are bounded between 0 and 1) are dichotomised at their median values, thus creating a binary variable IV.

The IV inequalities can jointly validate the ER and the independence assumptions (Balke and Pearl, 1997). For a binary instrumental variable and a binary outcome, an IV needs to meet IV inequalities (see Equation 1 in chapter 2). If these inequalities do not hold, ER and independence assumption are likely to be violated. Standardised mean difference (SMD) is used to examine balance in observed covariates (see Equation 6 and Equation 7). SMD less than 0.1 is considered insignificant imbalance (Austin, 2009).

$$SMD = \frac{\overline{X_1} - \overline{X_2}}{\sqrt[2]{(S_1^2 + S_2^2)/2}}$$

Equation 6. $\overline{X_1}$, $\overline{X_2}$: sample mean of treated and controlled group(defined by the treatment or the IV).S_1, S_2: sample variance of the two treatment groups (Flury and Riedwyl, 1986).

If the variable is dichotomous, the SMD is expressed as Equation 7.

$$SMD = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{[\hat{p}_1(1 - \hat{p}_1) + \hat{p}_2(1 - \hat{p}_2)]/2}}$$

Equation 7. p_1 and p_2 are prevalence of treated and control groups respectively (Austin, 2009)

Monotonicity assumption is validated using the monotonicity inequality proposed by Balke and Pearl (Balke and Pearl, 1997) (see Equation 2).

4.5.3 Study population

The Scottish national prescribing system (PIS) is a national individual-level routine data set which contains prescribing records for the 5.3 million residents in Scotland (Alvarez-Madrazo et al., 2016). In Scottish Morbidity record (SMR01), alcohol intoxication and harmful use (AIH) hospitalised patients are defined using International Classification of Diseases 10th Revision (ICD-10) coding of F10.0 or F10.1. There are six diagnosis fields in SMR01 for reason of hospitalisation and death, including one primary diagnosis and five secondary diagnoses. The follow-up time starts on 1st January 2010, ends on 31st December 2019; a period of 10 years. Patients with an AIH hospitalisation within 1 year prior to the first prescription of diazepam or chlordiazepoxide hydrochloride represent the cohort under study.

The covariates in this study are sex, age, Scottish Index of Multiple Deprivation (SIMD 2016), defined daily dosage (ddd), length of stay of incident AIH hospitalisation prior to the prescription, year of prescription. Considering the patients comorbidity may have changed across time, Charlson Comorbidity Index (CCI) within 1 year and 10 years are included as covariates in the regression models. Patients with missing data for the SIMD variable were removed (7.3%). The outcomes, measured from 1st January 2010 to 31st December 2019 are: 1) AIH rehospitalisation within 1 year (binary outcome); 2) time to AIH rehospitalisation (time-to-event outcome); 3) AIH death (binary outcome); and 4) time to AIH death (time-to-event outcome). The AIH rehospitalisation was defined as F10.0 or F10.1 in main condition field. The AIH related death is defined as F10.0 or F10.1 in main condition field or in the five other condition fields.

4.5.4 Statistical methods

The most commonly used IV estimation method is two-stage least squares (2SLS) where the first stage is an ordinary least square regression model built on the treatment and instrumental variable. The second stage of regression uses the predicted results from the first stage as well as the covariates included in the first stage regression to predict the outcome (Angrist and Imbens, 1995). When the outcome is binary, using 2SLS means treating the binary variable as if it is a numerical variable which may cause inconsistent estimates (Kuo et al., 2012). However, many researchers have compared the 2SLS with other estimation models which are more suitable for binary variables and concluded that 2SLS is generally unbiased (Ionescu-Ittu et al., 2009, Chapman and Brooks, 2016, Zhang et al., 2018). OLS and propensity score are selected as the conventional multivariable regression methods to compare with IV inference using 2SLS.

For the outcomes of time until AIH rehospitalisation or AIH death, I conducted conventional multivariable adjusted Cox proportional hazards regression and IV-based Cox proportional hazards regression (Tchetgen et al., 2015). The conventional Cox proportional hazards regressions are conducted by 'coxph' function from 'survival' package in R and include modelling of the covariates: sex, age, Scottish Index of Multiple Deprivation (simd 2016), defined daily dosage (ddd), length of stay of incident AIH hospitalisation prior to the prescription, year of prescription, comorbidity (CCI). The same group of covariates are included in the both stages of IV-based Cox regression model. The IV-based Cox proportional hazards regression is conducted using control function approach by including the residual from the first stage of regression as an additional covariate in the second stage regression (Terza et al., 2008a, Tchetgen et al., 2015). The statistical significance level was set at 5%. All statistical analysis was done inside the R studio (version 4.1.1).

4.5.5 Sensitivity analysis for unmeasured confounding

Sensitivity analysis for potential unmeasured confounding was conducted. This approach utilises a measured covariate as a benchmark covariate to quantify the magnitude of unmeasured confounding effect and shows the corresponding treatment effect under those times of the effect of benchmark covariates. If the estimation of treatment effect is not affected to a substantive extent, then the population under study is unlikely to be sensitive to unmeasured confounding (Cinelli and Hazlett, 2020). This approach was carried out using the R function 'sensemakr'(Cinelli et al., 2021). A corresponding sensitivity analysis was also applied for survival data using the R function 'survSens' (Huang et al., 2020). This approach assumes the unmeasured confounder is a 50/50 distributed binary variable and uses two sensitivity parameters to measure, 1) the association between unmeasured confounder and treatment; and 2) the association between unmeasured confounder and time-to-event outcome (Huang et al., 2020). The Wu-Hausman test was also conducted. If the null hypothesis is rejected indicating possible unmeasured confounding or treatment effect heterogeneity (Hausman, 1978).

4.6 Results

4.6.1 Descriptive statistics

It can be seen from Table 8, that diazepam is prescribed at a level more than twice than that for chlordiazepoxide hydrochloride in this AIH hospitalisation cohort. Patients from the most socio-economically deprived group (SIMD=1), based on grouping into ‘tenths’ using decile cut-points, account for the highest proportion (almost 24%). Note that there is a smaller percentage of female patients than male patients, especially for chlordiazepoxide hydrochloride prescription. On average, the length of hospitalisation for the diazepam group is longer than for the chlordiazepoxide hydrochloride group.

	CHLORDIAZEPOXIDE HYDROCHLORIDE (N=993)	DIAZEPAM (N=1897)	Overall (N=2890)
Age (years)			
Mean (SD)	46.6 (12.1)	46.3 (13.6)	46.4 (13.1)
Median [Min, Max]	46.0 [18.0, 83.0]	46.0 [14.0, 93.0]	46.0 [14.0, 93.0]
Gender			
Female	318 (32.0%)	835 (44.0%)	1153 (39.9%)
Male	675 (68.0%)	1062 (56.0%)	1737 (60.1%)
Defined Daily Dosage (mg)			
Mean (SD)	9.89 (6.78)	14.8 (20.1)	13.1 (16.9)
Median [Min, Max]	9.33 [0.333, 80.0]	7.50 [0.400, 168]	8.33 [0.333, 168]
Missing	0 (0%)	1 (0.1%)	1 (0.0%)
Scottish Index of multiple deprivation (2016)			
1 (most deprived)	267 (26.9%)	419 (22.1%)	686 (23.7%)
2	156 (15.7%)	292 (15.4%)	448 (15.5%)
3	149 (15.0%)	259 (13.7%)	408 (14.1%)
4	115 (11.6%)	207 (10.9%)	322 (11.1%)
5	92 (9.3%)	199 (10.5%)	291 (10.1%)
6	65 (6.5%)	155 (8.2%)	220 (7.6%)
7	48 (4.8%)	112 (5.9%)	160 (5.5%)
8	39 (3.9%)	92 (4.8%)	131 (4.5%)
9	36 (3.6%)	95 (5.0%)	131 (4.5%)
10	18 (1.8%)	55 (2.9%)	73 (2.5%)
NA	8 (0.8%)	12 (0.6%)	20 (0.7%)
Charlson comorbidity index 1 year			
0	637 (64.1%)	1199 (63.2%)	1836 (63.5%)
1	356 (35.9%)	698 (36.8%)	1054 (36.5%)
Charlson comorbidity index 10 year			
0	785 (79.1%)	1472 (77.6%)	2257 (78.1%)
1	208 (20.9%)	425 (22.4%)	633 (21.9%)
Prescription year			
2010	124 (12.5%)	171 (9.0%)	295 (10.2%)
2011	178 (17.9%)	211 (11.1%)	389 (13.5%)
2012	140 (14.1%)	200 (10.5%)	340 (11.8%)
2013	110 (11.1%)	197 (10.4%)	307 (10.6%)
2014	132 (13.3%)	235 (12.4%)	367 (12.7%)
2015	88 (8.9%)	187 (9.9%)	275 (9.5%)
2016	82 (8.3%)	199 (10.5%)	281 (9.7%)
2017	74 (7.5%)	218 (11.5%)	292 (10.1%)
2018	55 (5.5%)	223 (11.8%)	278 (9.6%)
2019	10 (1.0%)	56 (3.0%)	66 (2.3%)
Length of staying in hospitals (days)			
Mean (SD)	1.55 (2.94)	1.68 (9.53)	1.64 (7.91)
Median [Min, Max]	1.00 [0, 42.0]	1.00 [0, 366]	1.00 [0, 366]

[2]

Table 8. Descriptive statistics of the study population

[2] Since about 64% of patients with CCI equals 0, I divided the CCI into binary variable to avoid small sample sizes in other values of CCI. Since there are less than 1% missing value of the SIMD value, I dropped the missing value in the statistical analysis section.

4.6.2 Validation of IV assumptions

4.6.2.1 Validation of the relevance assumption

As it shows in Table 9, the F-statistics of the numerical instruments built using the different formulations can be considered as strong as they exceed the rule-of-thumb F-statistics of 10 (Stock et al., 2002). When using dichotomised PPP IVs formulations (based on medians), the F-statistics and AUC were similar. AUC over 0.70 indicates a relatively strong association between IV and treatment (Vanagas, 2004).

Proxies for the physician's prescribing preference (Definition of prior n IV see Figure 3 above)	F-statistics of numerical PPP IV	AUC of numerical PPP IV	F-statistics of dichotomised PPP IV	AUC of dichotomised PPP IV
Prior 1	10.531	0.667	10.531	0.667
Prior 2	36.064	0.673	35.289	0.673
Prior 3	32.726	0.673	35.793	0.673
Prior 4	44.257	0.674	43.463	0.673
Prior 5	54.528	0.677	61.266	0.677
Prior 6	66.338	0.680	73.385	0.681
Prior 7	76.813	0.683	86.304	0.684
Prior 8	90.482	0.687	96.018	0.687
Prior 9	88.466	0.688	93.238	0.687
Prior 10	91.134	0.688	97.695	0.688
Prior 11	93.332	0.689	99.679	0.689
Prior 12	102.771	0.690	105.653	0.691
Prior 13	113.420	0.691	122.918	0.696
Prior 14	122.605	0.694	129.949	0.698
Prior 15	128.292	0.697	133.131	0.698
Prior 16	134.546	0.699	147.946	0.702
Prior 17	137.443	0.701	150.065	0.703
Prior 18	142.563	0.702	142.967	0.702
Prior 19	150.261	0.703	142.089	0.702
Prior 20	144.631	0.702	144.183	0.702
Proportion	150.876	0.706	158.128	0.705

Table 9. Validation of the relevance assumption using F-statistics and AUC.

4.6.2.2 Validation of monotonicity assumption

The results from validation of the monotonicity assumption using the monotonicity inequality are shown in Table S3 in supplementary material. The details of monotonicity inequality had been described in Chapter 2. When the outcome is death, the monotonicity assumption is likely to be violated. However, the proportional IV is shown as valid IV for both types of outcomes. For exploration, I still include the other types of PPP IV in further analysis.

4.6.2.3 Validation of exclusion inclusion and independence assumptions

SMD for covariates based on IV are less than 0.1 indicating IV is valid in balancing covariates. There are noticeable deductions of SMD in terms of gender and dosage when they are more than 0.1 based on treatment and less than 0.1 based on IV. However, apart from that, the reduction of covariate balance is not visible as the SMD on the treatment level are naturally less than 0.1 (see Figure S1 and Figure S2 in supplementary material for Chapter 4).

4.6.3 Comparison of estimates of the treatment effect

Shown in Figure 4, diazepam is associated with lower risk of rehospitalisation and death than chlordiazepoxide hydrochloride. Further, the 2SLS estimates are further away from the null hypothesis compared with the OLS estimates and PS estimate which suggests an association (potentially causal) is detected using IV methods. Almost all the 2SLS estimates are not statistically significant due to large imprecision in comparison to multivariable regression and propensity score results. It also indicates that the estimated treatment effect of dichotomised PPP IV is quite different from that of non-dichotomised PPP IV. Generally, the 2SLS estimates from dichotomised PPP IV are closer to the null estimates.

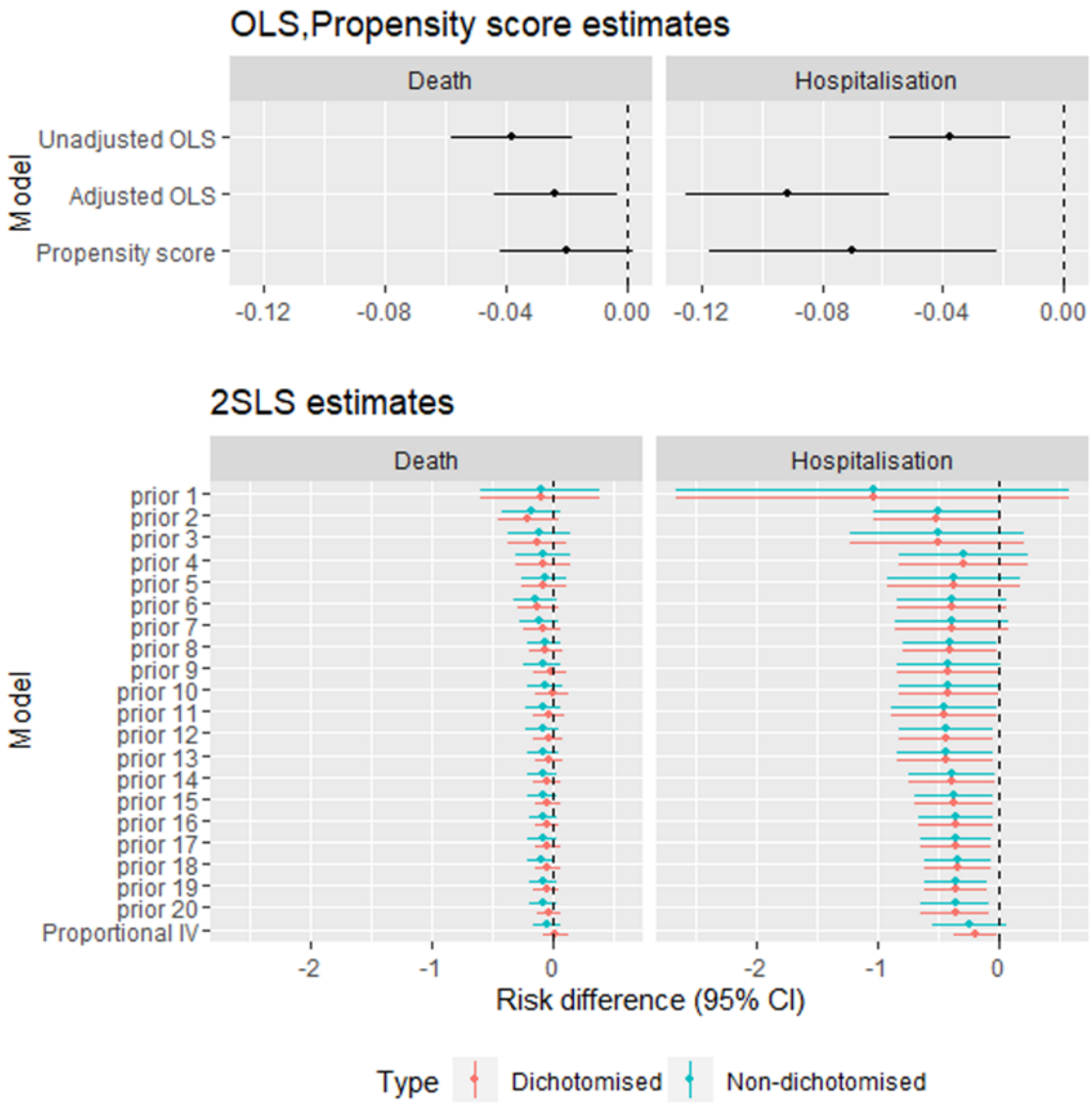


Figure 4. OLS, propensity score and 2SLS estimates.

4.3.3.1 Sensitivity analysis for unmeasured confounding

Figure S3 and Figure S4 in supplementary material shows that the potential unmeasured confounder needs to explain at least 4.78% and 4.53% of residual variance of treatment and outcome to make the estimated treatment effect across the null hypothesis.

4.6.4 Survival analysis

As can be seen in Figure 5, results from conventional Cox proportional regression indicate that diazepam is associated with lower relative risk of AIH rehospitalization and AIH death. IV results also show a protective effect from diazepam but is not statistically significant due to much wider confidence intervals indicating a lack of precision. When the outcome is rehospitalisation, IV results generate similar estimates of HR. However, in the case of death as outcome, the effect sizes of IV estimates are larger than that from conventional Cox proportional regression (explanation in section 4.6). There is no noticeable difference between the numerical IV (non-dichotomised) results and dichotomised IV results as all of them are statistically insignificant.

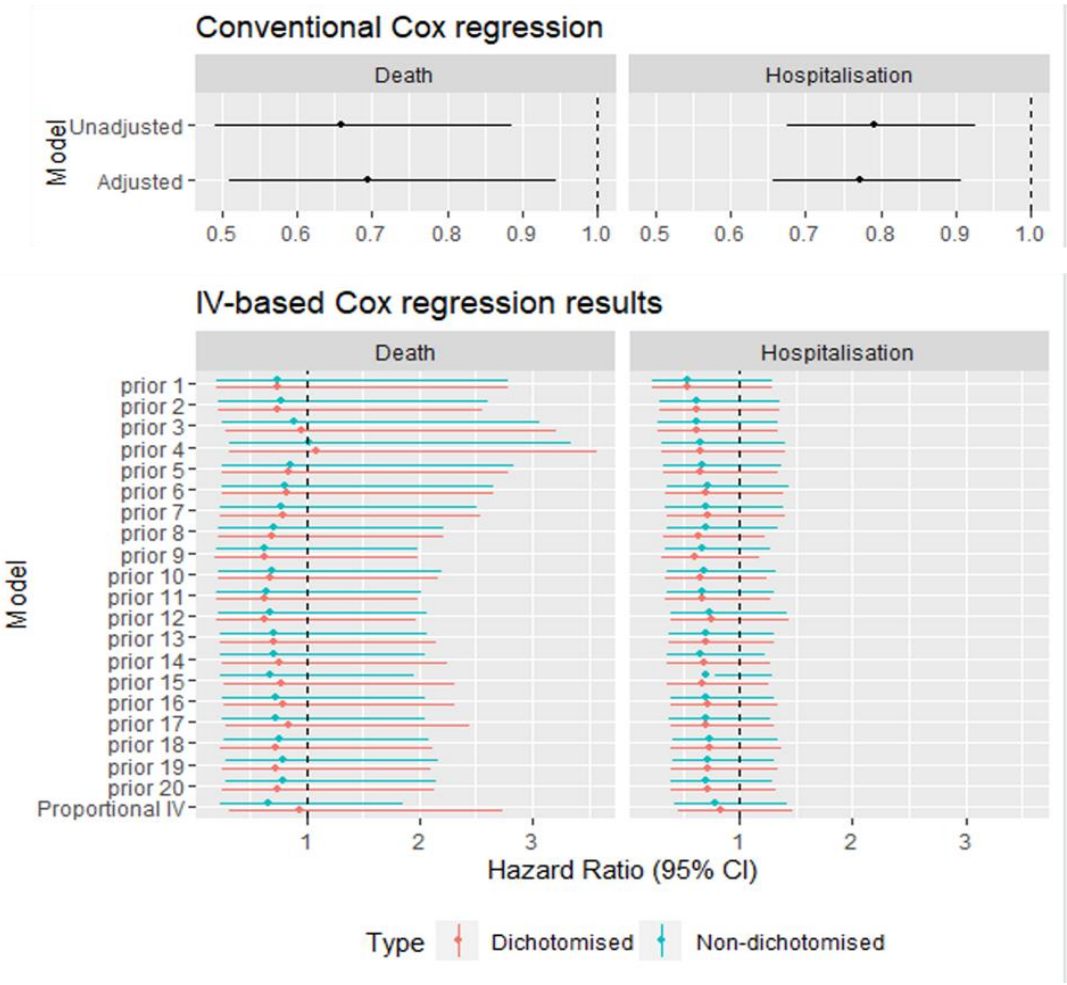
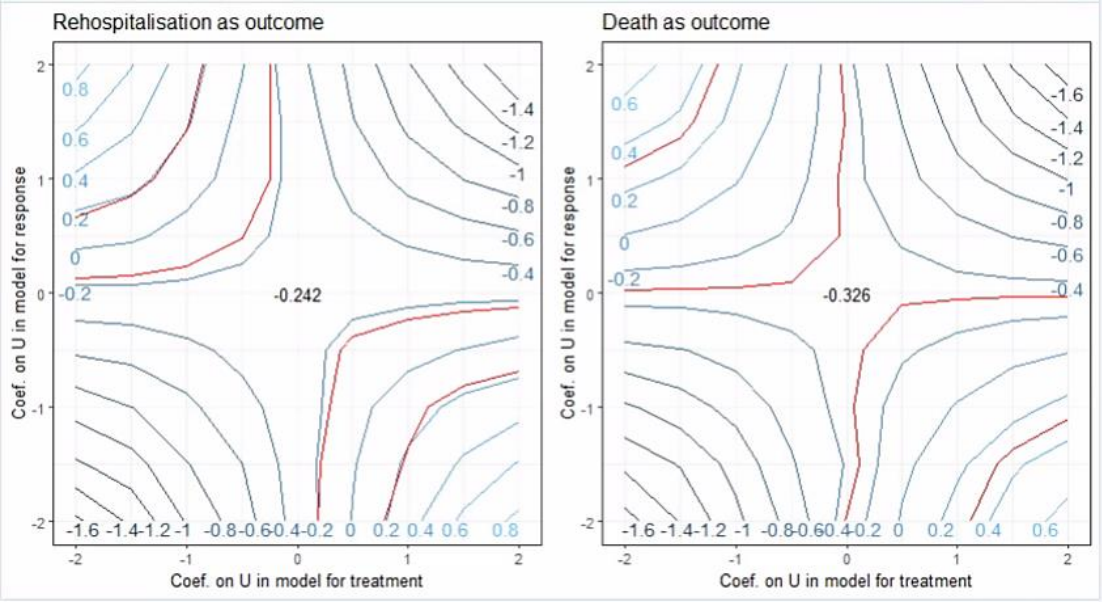


Figure 5. Conventional Cox regression and IV-based Cox regression results

4.3.4.1 Sensitivity analysis for unmeasured confounding (survival analysis)

It can be seen from Figure 6, a relatively strong association between outcome and treatment (for example, (-0.5, 1)) can lead to nonsignificant estimated treatment effect (inside the red curve area) when AIH rehospitalisation being the outcome. In terms of AIH death being the outcome, this association can be weaker, for example, (-0.5.0.2) in the upper left corner can make the estimated treatment effect be inside of the red curve area. The unmeasured confounding is more likely in this case. The results from the Wu-Hausman test were presented in the Table S3. The rejection of null hypothesis indicates the possibility of unmeasured confounding.



[3]

Figure 6. Sensitivity analysis results for death as outcome, the blue contour shows that sensitivity parameters corresponding estimated treatment effect.

[3] The red curve is the t-statistics equals 1.96. U on the axis represent the assumed unmeasured confounders. Area that surrounded by the red curve indicate insignificant treatment effects. Value at the centre is the value of estimated treatment effect ignoring any confounding.

4.7 Discussion

The main objective of this study is to investigate the performance of IV method with PPP as a potential IV. Before conducting IV approach, a sensitivity analysis of unmeasured confounding indicates that there is a possibility that the unmeasured confounding exists as a potential issue. The results from Figure S3 and Figure S4 from supplementary material the sensitivity analysis indicate that the unmeasured confounding is more likely to occur when the outcome is rehospitalisation in the OLS, and for outcome being AIH death in the Cox regression model. When unmeasured confounding is likely and the IV is valid, the IV approach results tend to be different with the conventional multivariable regression as IV is meant to solve the unmeasured confounding issue. This is reflected in Figure 4 and Figure 5. In Figure 4, the difference in estimated treatment effect from OLS and 2SLS is larger for the AIH rehospitalisation as the outcome. While in Figure 5, Cox regression and IV Cox regression model shows a greater difference in the estimated HRs of AIH death than that in the AIH rehospitalisation.

However, none of the IV results is statistically significant. This is partly because the sample size is moderate and the nature of two-stage methods (Ionescu-Ittu et al., 2009).This can be explained by Equation 8 below. The smaller sample size, the larger variance of 2SLS estimate. In Figure 4, it is noticeable that the confidence interval shrinks as the strength of IV increases which can also be explained by the Equation 8 in which the size of variance of 2SLS estimate will decrease as the association between the treatment and IV gets stronger (Martens et al., 2006).

$$var\left(\hat{\beta}_n^{IV}\right) = \frac{\sigma_{Y,X}^2}{n\,\sigma_X^2 \rho_{X,Z}^2}$$ [4]

Equation 8. Variance of 2SLS estimate.

[4] $\sigma_{Y,X}^2$: The residual variance of the outcome after adjusting the treatment (X); σ_X^2: The variance of the treatment; $\rho_{X,Z}^2$: The correlation between the treatment(X) and the instrumental variable (Z). n: sample size.

In my results, the proportional IV turns out to be the strongest IV and generates the smallest parameter estimate. Since the IV estimand can be nonparametrically presented as $Cov(Y,Z)/Cov(X,Z)$ or (E[Y=1|Z=1]-E[Y=0|Z=0])/E[X=1|Z=1]-E[X=1|Z=0] as I mentioned in Chapter 2, I estimated the values of IV estimand by using such formula (See Table S4 in supplementary material). The values of estimated IV estimand shows that the proportion IV has the lowest value of estimated IV estimand which indicate the proportion IV generates the smallest estimated treatment effect.

Given the monotonicity assumption and independence assumption are validated using dichotomised PPP IV, IV results from dichotomised PPP IV and non-dichotomised PPP IV are compared. The dichotomised PPP IV is preferred in many CERs because it is easier to validate IV assumptions (Baiocchi et al., 2014) and more straightforward to understand the estimated LATE. Ideally, one should present the results from both forms of PPP IV. Angrist and colleagues indicate that stronger IVs suffer less from the bias caused by the violation of monotonicity so as the bias from violation of other assumptions (Angrist et al., 1996). By applying stronger IV, one can reduce the harm from violating the monotonicity assumption. In general, this chapter provides evidence for favouring stronger IVs.

There are major differences observed in the absolute risk (see Figure 4) and relative risks (see Figure 5). In Figure 4, OLS and 2SLS generate small size of treatment effect on the absolute scale. However, in Figure 5, Cox proportional regression indicates diazepam users were 22% less relative risk of AIH rehospitalisation and 30% less relative risk of AIH death in comparison with chlordiazepoxide. In these cases, I tend to agree with the survival analysis approach as it taking censoring into account. However, care has to be taken in not overinterpreting HR results far from the null when the absolute risk of the outcome under study is small.

In the Cox regression and OLS regression results, my findings support that diazepam is a better choice for treating AIH patients than chlordiazepoxide hydrochloride in terms of preventing AIH rehospitalisation and AIH death. This conclusion echoes with the results from Jauhar et al. that diazepam is superior to chlordiazepoxide hydrochloride (Jauhar and Anderson, 2000) and adds real-world evidence of the treatment effects on AIH. Compared with AWS, AIH is normally overlooked, and not having a specific recommended pharmacological treatment (Jung and Namkoong, 2014). At the same time, acute alcohol intoxication may predict AWS as the treatment for AIH can also influence AWS outcomes (Mirijello et al., 2015). AIH and AWS have overlaps and need to be more investigated in unison. In this observational study, I chose to investigate AIH over AWS due to the small sample sizes of AWS patient cohorts (and consequent events). Despite the potential caveat about the unmeasured confounding, there is no detectable difference between the treatment estimates between the approaches. In this chapter, the IV results tend to be statistically insignificant and the results from the conventional approaches should be preferred.

4.8 Strengths and Limitations

Findings from this chapter suggest a gap between the NICE guideline and the commonly prescribed prescription in real-life settings. I discovered that many patients with AIH hospitalisation history were prescribed with diazepam and chlordiazepoxide hydrochloride and found significant difference in their effect on outcomes. Besides, findings also indicate that the disease severity is a potential source of unmeasured confounding. Although I have added CCI which draw from the hospitalisation history of the patients, I do not know how 'sick' a patient is relative to each other. From the descriptive statistics (see Table 8), the average of length of staying in hospital is longer for patients who have been prescribed with diazepam than patients with chlordiazepoxide. IV results tend to be more away from null indicating patients with higher expected benefits (possibly be sicker patients) tend not get diazepam.

The characteristics of the preference-based IV is a source of limitation as it cannot be measured directly and needs proxies. The proxy for the IV is treated as non-causal IV which is hard to identify and leads to issues in the validation of the monotonicity assumption (Labrecque and Swanson, 2018). Another limitation of this study is that some assumptions cannot be validated empirically. For example, if one wants to validate the monotonicity assumption, one can design a questionnaire survey for the physicians and ask them about their prescribing preference between two medications and their treatment for groups of patients with specific set of characteristics. If the prescribing preference and the prescriptions they prescribe to the patients match, one can assume the monotonicity assumption holds. According to our results, PPP IV that accounts for longer length of prescription history tends to reflect the prescribing preference in a more precise way. The F-statistics, OR and AUC from the regression on exposure and IV show that the strength of IV increases as longer prescription history being accounted. The strength of instrument variable is essential for the performance of IV methods. With a strong enough IV and other assumptions hold, the IV estimates are reliable to be compared with the conventional results.

4.9 Conclusion

In large real-world health data sets, diazepam has a stronger protective effect against AIH re-hospitalisation and AIH death than chlordiazepoxide hydrochloride. Where possible, the instrumental variable method should be a complementary analysis in comparative effectiveness research using observational data when unmeasured confounding is of major concern (which it usually is).

Chapter 5. Epidemiology of pharmacological treatments for alcohol dependence in the UK: evidence from primary and secondary healthcare data

5.1 Publication and author contribution details

This manuscript had been submitted to Drug and Alcohol Review for publication and currently under peer review.

I jointly contributed with Francesco Manca and Jim Lewsey in drafting overall manuscript. I contributed independently to trends and rates using CPRD data (section 5.6.1.1), inequality analysis using CPRD data (section 5.6.2.1), IV analysis using CPRD data and PIS data (section 5.6.3), and the interpretation of IV results. Francesco Manca contributed independently to trends and rates using PIS data (section 5.6.1.2), inequality analysis from PIS data (section 5.6.2.2), non-IV statistical analysis using PIS data (section 5.6.3), and most of the initial drafting of this manuscript.

5.2 Data used in this chapter.

CPRD data used in this chapter was applied to CRPD under the application number #20_000126. The application document is attached in supplementary material for Chapter 5. PIS data (Scottish cohort) in this chapter was applied under the same application procedure as Chapter 4. The research objective covered in this chapter requires the prescription record (for alcohol use disorder, both CRPD AURUM and CPRD GOLD), and the outcome variables including hospitalisation records and death records. For that reason, HES Admitted Patient Care was applied. The CPRD linkage request form is attached in the supplementary material.

The alcohol dependence cohort is identified using the read codes according to Thompson's work (Thompson et al., 2017) from the primary care data in the CPRD GOLD and AURUM. The read code list is attached in Table S8 in supplementary material for Chapter 5. Data tables and variables that are included in statistical analysis are listed below. The description of the variables can be found on the website of CPRD (https://cprd.com/sites/default/files/2022-02/Data_Dictionary_HES_APC.pdf).

Tables	**Variables**
GOLD/Patient	patid gender Yob frd crd
Gold/Staff	Staffid Role
Gold/ therapy	patid eventdate prodcode staffed
Gold/clinical	patid eventdate Medcode
Aurum/ Patient	patid gender yob
Aurum/ staff	staffid jobcatid
Aurum /Observation	patid enterdate Medcodeid
Aurum/DrugIssue	patid Issuedate Staffid Prodcodeid
hes_diagnosis_hosp in Gold and Aurum	patid admidate dischargedate ICD ICDx
practice_imd	pracid country imd
Linkage Source (linkage_eligibility.txt)	patid pracid

5.3 Abstract

Background

Effective pharmacological treatments for alcohol-dependence are under-prescribed in the UK. We assessed, for both primary and secondary healthcare, the prevalence of such prescribing and the extent of any inequalities. Further, we compared the effectiveness of the two most prescribed drugs (acamprosate and disulfiram) and assessed whether there is inequality in prescribing either of them.

Methods

We used two healthcare databases: general practice (Clinical Practice Research Datalink) for England and hospitalisations and dispensed prescriptions for Scotland. Logistic regression was used to assess the odds of receiving any alcohol-dependence prescription in both primary and secondary healthcare, and the comparative odds of receiving acamprosate or disulfiram. Comparative effectiveness was assessed using time-to-event modelling.

Results

Only 2-4% of patients with alcohol-dependent diagnoses in primary healthcare received alcohol-dependence prescriptions within 60 days after diagnosis. Inequalities in prescribing existed, especially across sex and age, with differences between healthcare setting. For example, being male had 36% lower odds of prescription (OR: 0.64, 95% CI: 0.60-0.68) in primary care and 12% (OR: 0.88, 95% CI:0.81-0.96) in secondary healthcare. Prescribing was similar by socio-economic deprivation in primary care, but lower among the most deprived after an incident secondary healthcare event. Disulfiram was superior to acamprosate in preventing alcohol-related hospitalisations (hazard ratio point estimates between 0.54-0.77). Acamprosate was more often prescribed for those residing in more deprived areas.

Conclusions

Prescribing of alcohol-dependence medications is low in the UK, with greater under-prescribing among men and less effective medications used more for those living in more socioeconomically deprived areas.

Key words: Alcohol dependence, comparative effectiveness, prescription inequality, routine health data, primary healthcare, secondary healthcare, physician prescribing preferences.

5.4 Introduction

Excessive alcohol use is related to a range of adverse health outcomes and causes societal as well as individual harm. Alcohol dependence, as defined by the National Institute for Health and Care Excellence (NICE), is “characterised by craving, tolerance, a preoccupation with alcohol and continued drinking in spite of harmful consequences” (National Collaborating Centre for Mental, 2011). Globally, in 2016, the estimated age-standardised prevalence of alcohol dependence was 1320.8 cases per 100,000 people (Collaborators, 2018). In the UK, between 1990 and 2013, the estimated rate of presentation to general practice with alcohol dependence was 171 and 76 per 100,000 male and female patients, respectively (Thompson et al., 2017).

Guidelines worldwide suggest pharmacological treatments for alcohol-dependent patients subsequent to detox and alongside psychosocial support, with specific drugs suggested based on patients' goals (reduction in consumption or total abstinence), comorbidities and the capability with potential side effects (Haber et al., 2009). In particular, concerning the UK, NICE clinical guidelines (CG115) recommend that for people with mild alcohol dependence a psychological intervention is offered, and for those with moderate/severe alcohol dependence these psychological interventions can be used in combination with the pharmacological intervention (National Collaborating Centre for Mental, 2011). In the UK, the two main medications prescribed for treating alcohol dependence are acamprosate and disulfiram. Acamprosate helps to maintain abstinence by restoring neurotransmitters affected by excessive alcohol use and can also contribute to managing alcohol cravings, but it is generally effective only in someone already sober (Mason and Heyser, 2010, Patel and Balasanova, 2021). In contrast, disulfiram causes unpleasant symptoms if alcohol is consumed, functioning as a deterrent to alcohol drinking. Due to its strong effects, manufacturers suggest that patients and their carers are counselled on the disulfiram-alcohol reaction and NICE advices monitoring patients in the initial phases of treatment (Excellence). In a small high-quality evidence base, two open label randomised trials compared disulfiram to acamprosate and showed disulfiram to be more effective in reducing alcohol intake, increasing the number of abstinence days and reducing risks of relapse (Laaksonen et al., 2008), and in increasing the percentage of abstinent patients and reducing risk of relapse (de Sousa and de Sousa, 2005). The pure use of clinical trials to inform guidelines may cause some criticism as they do not usually look at long-term outcomes which are relevant especially regarding addiction. Therefore, stronger evidence from real-world effectiveness based on larger populations and with less time constraint is crucial. To date, only a small observational study that directly compared the two drugs, showing that disulfiram has a longer duration of time to alcohol relapse and higher cumulative abstinence compared to acamprosate (Diehl et al., 2010).

Despite evidence on the effectiveness and their inclusion in clinical guidelines, pharmacological intervention for treating alcohol dependence is underutilised in clinical practice with many patients not getting prescribed the specialised treatment (Antonelli et al., 2022). When there is evidence of underutilisation, it is important to understand whether this is caused, at least in part, by some groups less likely to get prescribed than others. If this happens, inequalities in health outcomes can be exacerbated if those less likely to get prescribed are those the most in need (i.e., more likely to experience severe alcohol dependence). Further, current evidence of under-prescriptions regards primary healthcare (Thompson et al., 2017), however, it is not clear whether other healthcare levels have similar characteristics.

Using two large routinely-collect healthcare datasets from the UK regarding primary and secondary healthcare, we aimed to assess the prescription levels of pharmacological treatment for alcohol dependence in different levels of healthcare and whether inequalities exist in prescribing levels by age, sex, and socio-economic deprivation. Further, we compare the real-world effectiveness of acamprosate and disulfiram to time to first alcohol-related hospitalisation, including using an instrumental variable approach to account for unobserved confounding. Lastly, we assess whether there is inequality in prescribing between these two medications.

5.5 Methods

5.5.1 Data sources

For all the analyses we used two separate datasets both referring to the UK. The first focusing on primary healthcare and the second on secondary healthcare. For the primary healthcare dataset, we utilised the English subset of the Clinical Practice Research Datalink (CPRD) (Herrett et al., 2015, Wolf et al., 2019), a UK wide dataset collecting data from a network of over 2,000 primary healthcare practices and broadly representative at the country level. The dataset identified patients in primary healthcare and linked them with prescriptions and future hospitalisations. For the secondary healthcare dataset, we utilised a Scottish dataset linking three nationwide administrative healthcare databases containing data from 2009 to 2019 regarding dispensed prescriptions from Scottish National Prescribing Information System (PIS), hospitalisations (SMR01) and deaths (National Records of Scotland). From this point when referring to the English primary healthcare dataset we will use 'Eng-CPRD', while for the Scottish secondary healthcare dataset we will use 'Scot-PIS'.

5.5.2 Pharmacological treatments

We evaluated the trend and inequalities in prescriptions for all medications in the guidelines with an exclusive indication for the treatment of moderate or severe alcohol dependence (National Collaborating Centre for Mental, 2011): acamprosate, disulfiram and nalmefene. However, nalmefene was rarely prescribed to individuals in our datasets and so we compared the effectiveness of the two most common prescriptions: acamprosate and disulfiram. We then ran a further analysis assessing inequality in prescriptions between these two drugs.

5.5.3 Statistical analyses

5.5.3.1 Trends and rates

For both levels of care, we assessed the trends of alcohol dependence prescriptions over time.

Primary healthcare Eng-CPRD

In Eng-CPRD, we checked the rate of first diagnosis for alcohol dependence defined by Read codes (Chisholm, 1990) ‘alcohol dependence and consequences of alcohol dependence' (Thompson et al., 2017) over time. We then observed the percentage of patients receiving prescriptions within 60 days after their diagnosis (looking at potential differences across socio-economic deprivation levels). The 60 days window had the purpose of associating the prescription with the diagnosis episode.

Secondary healthcare Scot-PIS

Similarly, in Scot-PIS we checked the rate of patients with a first hospitalisation of ‘mental and behavioural disorders due to alcohol’ (ICD F10.x, main diagnostic position) within the national population and then the percentage of individuals receiving alcohol dependence prescriptions within 60 days after discharge.

5.5.3.2 Inequality

Primary healthcare Eng-CPRD

The cohort was defined between January 2010 and December 2019 and identified all patients with a first diagnosis of alcohol dependence in primary healthcare (see above for inclusion criteria in Eng-CPRD), we excluded patients with previous hospitalisation for AUD. We determined whether patients received prescriptions within 60 days from their diagnosis. We repeated the same analysis on prescriptions received any time after the diagnosis. Logistic regression was used to assess whether age, sex and socio-economic deprivation of the practice were associated with the odds of prescriptions for alcohol dependence. We adjusted the regression by year as every calendar year the dataset changes the overall population at risk with new patients and practices subscribing. Whenever the relationship between covariates and the dependent variable was not linear (e.g., for age), restricted cubic splines (Gauthier et al., 2020) were used. The same analysis was run for the two most prescribed drugs to explore whether different patient characteristics might lead to different drug prescribing.

Secondary healthcare Scot-PIS

We identified a cohort between January 2010 and March 2019 with a first hospitalisation of alcohol use disorder (AUD) diagnoses in the main diagnostic position (see above for inclusion criteria in Scot-PIS) screening back for 10 years to avoid previous alcohol-related hospitalisation. We ran a similar analysis to Eng-CPRD, determining inequality in prescriptions for alcohol dependence within 60 days of hospital discharge and repeating the analysis with medications received at any time after AUD hospitalisations. The main differences between the two datasets were that in Scot-PIS we were able to adjust the model for comorbidities (measured through Charlson comorbidity score (D'Hoore et al., 1996) and previous hospitalisation related to mental health) and for alcohol dependence prescriptions taken before hospitalisation. Further, the socio-economic deprivation in Scot-PIS was rereferred to the patient. Using the same Scot-PIS cohort, we also did an additional analysis using as the dependent variable obtaining a prescription before the hospitalisation, aimed to assess imbalances of prescriptions in preventing patients to be hospitalised (see Table S5 in supplementary material for Chapter 5).

5.5.3.3 Comparative effectiveness

In both datasets, we defined the cohort for this analysis in the same fashion. We identified patients with a first prescription of acamprosate or disulfiram without any previous hospitalisation for F10.x in the previous 5 years (the reduced time compared to the previous analyses was due to the mismatch in backwards data between datasets). The outcome under study was time to first hospitalisation for F10.x after prescription. We assessed time to first hospitalisation using three different approaches: Cox regression adjusted for covariates (age, sex, socio-economic deprivation), covariates used n propensity scores (inverse probability weight using the same covariates) and an PPP IV (Brookhart et al., 2006).

For PPP IV, we implemented 2SRI which provide consistent estimators in non-linear models (Terza et al., 2008a). The instrument we used in our 2SRI-Cox model is the proportion of acamprosate prescribed by a particular physician in the last 10 prescriptions. Given the high number of new practices added every year in Eng-CPRD, 60% of general practitioners with at least one prescription with an indication for alcohol dependence did not have a history of 10 previous prescriptions for alcohol dependence recorded in the dataset. For this reason, we used multiple imputation by chained equation to build the PPP instrument (Janssen et al., 2010). While the first two approaches controlled for measured confounding by indication, the IV approach accounted for potential unmeasured confounding assuming all the assumptions are met. The baseline demographics of the cohorts identified for the inequality and comparative effectiveness analyses are presented in Table 10.

		N.	Age, years mean (sd)	Sex (male)	Charlson Comorbidity Index		Previous hospitalisations for mental health	Socio-economic multiple deprivation in quintiles				
					0	>=1		1st (most deprived)	2nd	3rd	4th	5th (least deprived)
Primary Care Eng-CPRD	Total (%)	69114	46.3 (13.2)	47834 (69%)	-	-	-	23839 (34%)	17995 (26%)	11138 (16%)	8551 (12%)	7591 (11%)
	Individuals receiving prescriptions	4581	43.8 (10.9)	2738 (60%)	-	-	-	1525 (33%)	1200 (26%)	764 (17%)	617 (13%)	475 (10%)
	Acamprosate*	3864	44.1 (10.9)	2298 (59%)	-	-	-	1313 (34%)	1028 (27%)	639 (17%)	511 (13%)	373 (10%)
	Disulfiram*	691	42.1 (10.5)	428 (62%)	-	-	-	203 (29%)	169 (24%)	122 (18%)	100 (14%)	97 (14%)
	Not receiving prescriptions	64533	46.5 (13.4)	45096 (70%)	-	-	-	22314 (35%)	16795 (26%)	10374 (16%)	7934 (12%)	7116 (11%)
Secondary Care Scot-PIS	Total (%)	19748	44.8 (18)	13463 (68%)	14049 (71%)	5699 (29)	6086 (31%)	5853 (30%)	4896 25%)	3925 (20%)	2964 (15%)	2110 (11%)
	Individuals receiving prescriptions	1240	46.1 (11)	841 (68%)	888 (72%)	352 (28%)	408 (33%)	275 (22%)	340 (27%)	278 (22%)	205 (17%)	142 (11%)
	Acamprosate*	840	46.5 (11)	541 (64%)	584 (70%)	256 (30%)	267 (32%)	197 (23%)	236 (28%)	183 (22%)	136 (16%)	88 (10%)
	Disulfiram*	349	45.3 (11)	222 (64%)	262 (75%)	87 (25%)	120 (34%)	67 (19%)	92 (26%)	80 (23%)	61 (17%)	49 (14%)
	Not receiving prescriptions	18508	44.2 (18)	12662 (68%)	13161 (71%)	5347 (29%)	5678 (31%)	5578 (30%)	4556 (25%)	3647 (20%)	2759 (15%)	1968 (11%)
Primary Care Eng-CPRD	Acamprosate	7033	45.4 (11)	4159 (59%)	-	-	-	706 (10%)	952 (14%)	1195 (17%)	1862 (26%)	2321 (33%)
	Disulfiram	1801	44.9 (11)	1100 (61%)	-	-	-	260 (14%)	315 (17%)	331 (18%)	423 (23%)	472 (26%)
Secondary Care Scot-PIS	Acamprosate	8016	44.9 (12)	4911 (61%)	-	-	-	2874 (36%)	1969 (25%)	1451 (18%)	1030 (13%)	692 (9%)
	Disulfiram	3223	43.2 (11)	2088 (65%)	-	-	-	964 (30%)	814 (25%)	626 (19%)	473 (15%)	346 (11%)

* The number of patients receiving acamprosate and disulfiram prescriptions does not equate to total number of patients as Nalmefene was also prescribed

Table 10. Characteristics of cohorts used in inequality and comparative effectiveness analyses.

5.6 Results

5.6.1 Trends and rates

5.6.1.1 Primary healthcare. Eng-CPRD

While the rate of alcohol dependence diagnosis decreased over the years in Eng-CPRD, there were not major differences in receiving prescriptions between most and least deprived groups (Figure 7,a).

5.6.1.2 Secondary healthcare. Scot-PIS

In contrast, in secondary healthcare the rate of AUD hospitalisation slightly increased over the years. The percentage of individuals with medications dispensed within 60 days after their first hospitalisations varied across socio-economic groups, with the least deprived groups receiving always more prescriptions after hospital discharge compared to the most deprived, except for the year 2016 (see Figure 7, b). In contrast, the relative difference in sex and age did not change between primary and secondary care over the years (see Figure S5, S6, S7, S8 in supplementary material for Chapter 5). Further, the percentage of prescriptions after hospitalisation is higher than in primary healthcare.

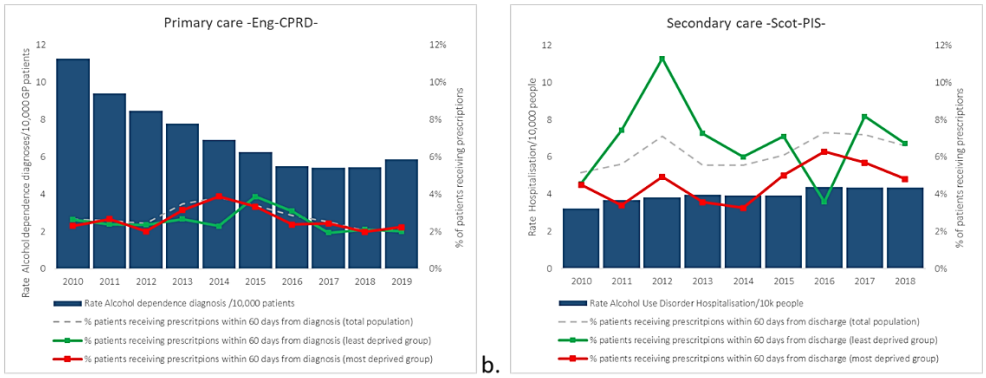


Figure 7. Trends and rates of alcohol dependence diagnosis (a) and alcohol use disorders hospitalisation (b) and percentage of such individuals receiving alcohol dependence prescriptions within 60 days of their diagnosis (a) or discharge date (b). In panel b 2019 was removed as data were only until March.

5.6.2 Inequality

5.6.2.1 Primary healthcare Eng-CPRD

Receiving prescriptions with indications for alcohol dependence after the first diagnosis of alcohol dependence in primary care was associated with sex (male less odds of receiving a prescription OR: 0.65, 95% CI: 0.60-0.71) and age (odds increasing until 41 years of age and then decreasing in older individuals (see Table S6 in supplementary materials for graphs showing modelled curvi-linear association with age). There were no substantial variations across deprivation groups. Comparing the two most prescribed drugs for alcohol dependence, odds of prescribing disulfiram decreased with age and increased in the least deprived group (OR: 1.55, 95%CI: 1.02-2.35). In both evaluations, odds of receiving prescriptions did not considerably vary by changing the time from diagnosis to prescription (Table 11. Inequality models of prescriptions after hospitalisation, before hospitalisation and inequality in prescriptions between the two most used alcohol dependence drugs

, model 1 vs 2 and model 3 vs 4).

Data			**Primary care - Eng-CPRD -**			**Secondary care - Scot-PIS -**		
Model5 (results in odds ratio)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sex (female reference)	0.654 (0.596-0.717)	0.643 (0.604-0.684)	1.087 (0.844-1.402)	1.102 (0.935-1.299)	0.877 (0.768-1.003)	0.878 (0.806-0.956)	1.002 (0.758-1.324)	1.095 (0.927-1.292)
Charlson Comorbidity Index (0 reference)								
>=1					0.922 (0.798-1.066)	0.914 (0.834-1.003)	0.840 (0.616-1.146)	0.768 (0.638-0.925)
Mental health comorbidity					0.919 (.801-1.054)	0.899 (0.824-0.981)	1.110 (0.832-1.481)	1.137 (0.960-1.346)
Simd= 1 as reference								
2	1.079 (0.959-.1.216)	1.069 (0.987-1.157)	0.965 (0.692-1.345)	1.096 (0.885-1.356)	1.407 (1.177-1.683)	1.046 (0.938-1.166)	1.198 (0.808-1.775)	1.419 (1.140-1.766)
3	1.227 (1.074-1.401)	1.102 (1.006-1.207)	0.990 (0.687-1.425)	1.231 (0.970-1.564)	1.624 (1.347-1.958)	1.110 (0.989-1.246)	1.414 (0.941-2.125)	1.286 (1.017-1.626)
4	1.222 (1.057-1.414)	1.169 (1.058-1.290)	1.174 (0.794-1.739)	1.193 (0.921-1.545)	1.540 (1.254-1.890)	1.068 (0.939-1.214)	1.411 (0.909-2.189)	1.469 (1.138-1.896)
5	0.992 (0.842-1.169)	0.993 (0.891-1.108)	1.550 (1.024-2.348)	1.659 (1.269-2.168)	1.454 (1.155-1.830)	1.115 (0.967-1.286)	1.917 (1.186-3.100)	1.782 (1.354-2.347)
Previous prescription in the previous 60 days†								
Any					23.42 (19.630-27.942)			
Acamprosate								0.510 (0.373-0.695)
Disulfiram								6.478 (4.818-8.712)
Previous prescription ever†								
Any						5.122 (4.664-5.625)		
Acamprosate							0.454 (0.300-0.687)	
Disulfiram							6.081 (4.067-9.072)	
Year 2020 as reference								
2011	0.975 (0.818-1.162)	1.017 (0.915-1.130)	0.719 (0.468-1.104)	0.808 (.629-1.039)				
2012	0.930 (0.774-1.117)	1.014 (0.909-1.131)	0.654 (0.413-1.037)	0.602 (0.455-0.794)				
2013	1.351 (1.143-1.608)	1.053 (0.942-1.176)	0.603 (0.391-0.930)	0.666 (0.505-0.878)				
2014	1.483 (1.258-1.775)	0.999 (0.889-1.122)	0.606 (0.391-0.942)	0.689 (0.514-0.924)				
2015	1.355 (1.129-1.627)	0.899 (0.794-1.109)	0.533 (0.329-0.861)	0.631 (0.458-0.869)				
2016	1.137 (0.931-1.391)	0.706 (0.613-0.813)	0.481 (0.276-0.841)	0.508 (0.343-0.752)				
2017	0.997 (0.808-1.231)	0.610 (0.526-0.709)	0.565 (0.323-0.990)	0.509 (0.393-0.883)				
2018	0.813 (0.648-1.020)	0.505 (0.431-0.593)	0.302 (0.144-0.631)	0.252 (0.140-0.453)				
2019	0.858 (0.690-1.066)	0.394 (0.332-0.467)	0.266 (0.128-0.556)	0.341 (0.192-0.604)				

As the relationship with age was not linear, we applied spline, see graph for interpretation of spline in Table S6 in the supplementary material.
†'Any' stands for 'any kind of prescription for alcohol dependence' and it is used in models referring to the overall inequality in secondary care (5) &(6), the variable is a dichotomous (0=no prescriptions and 1=prescriptions). Models (7) & (8), comparing acamprosate and disulfiram, specify the kind of prescription received before hospitalisation in a trichotomous variable (0= no previous prescriptions, 1=previous acamprosate prescriptions, 2=previous disulfiram prescriptions) where the reference is 'no previous prescriptions.

Table 11. Inequality models of prescriptions after hospitalisation, before hospitalisation and inequality in prescriptions between the two most used alcohol dependence drugs

[5] Model (1) Any alcohol dependence prescription after 60 days from diagnosis ; (2) Any alcohol dependence prescription ever from diagnosis;(3)Acamprosate vs disulfiram prescriptions within 60 days of diagnosis-acamprosate as the reference; (4) Acamprosate vs disulfiram prescriptions ever after diagnosis -acamprosate as the reference; (5) Any alcohol dependence prescription after 60 days from hospitalisation; (6) Any alcohol dependence prescription ever from hospitalisation ; (7) Acamprosate vs disulfiram prescriptions within 60 days of hospitalisation-acamprosate as reference; (8) Acamprosate vs disulfiram prescriptions ever after hospitalisation -acamprosate as reference-

5.6.2.2 Secondary healthcare Scot-PIS

Similarly, receiving prescriptions for alcohol dependence after 60 days from an AUD hospitalisation was associated with age. Socio-economic deprivation was also a factor associated with odds of receiving prescriptions after a secondary healthcare episode: living in least deprived areas was significantly associated with an increase in odds of receiving prescriptions of at least 41% (OR: 1.41, 95% CI:1.18-1.68 -values for the second most deprived quintile-) (see Table 11. Inequality models of prescriptions after hospitalisation, before hospitalisation and inequality in prescriptions between the two most used alcohol dependence drugs

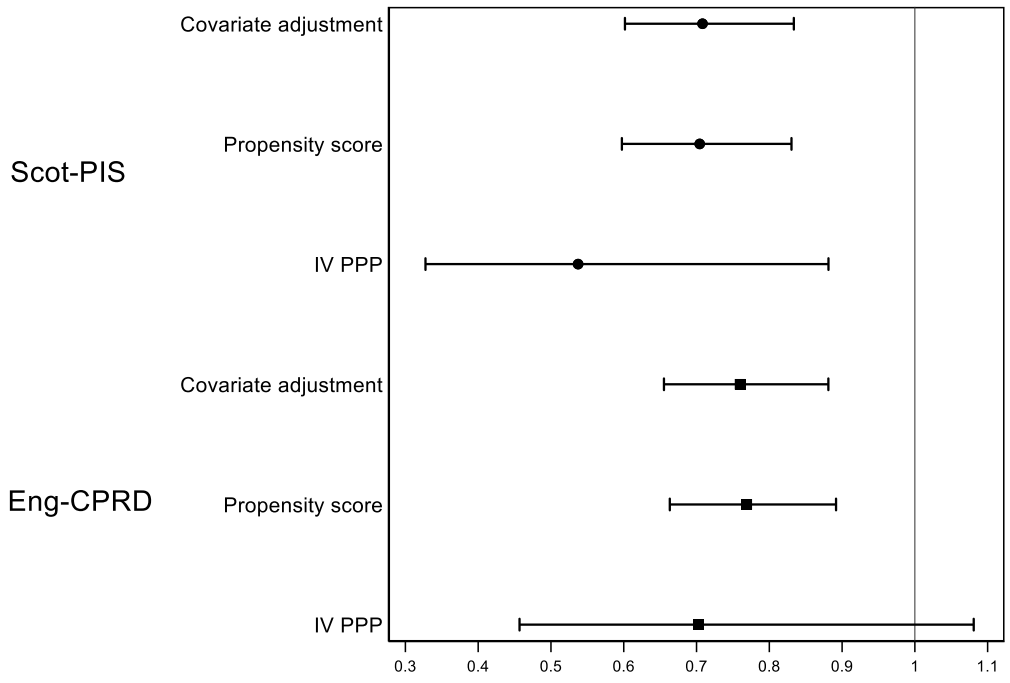
, column 5). Lastly, receiving prescriptions prior to hospitalisation was associated with a 23-fold increase (OR: 23.42, 95% CI: 19.63-27.94) in the odds of receiving prescriptions later. Other factors such as mental health comorbidities and sex did not have strong associations with prescriptions just after being discharged but became more precise (and statistically significant, p<0.01) when we did not include the 60 days constraint after hospitalisation (see Table 11. Inequality models of prescriptions after hospitalisation, before hospitalisation and inequality in prescriptions between the two most used alcohol dependence drugs

in sum, column 6). When we analysed odds of getting prescriptions before hospitalisations (see Table 7 in supplementary material), comorbidities (and in particular mental health comorbidities) were associated with an increment in the odds (OR=1.32, 95%CI:1.20-1.44), in contrast they were associated with reduction in the odds of getting prescriptions after hospitalisation in the long term (OR=0.90, 95%CI: 0.82-0.98). Similarly, to Eng-CPRD, we found that the odds of receiving disulfiram instead of acamprosate were associated with deprivation but also with the kind of medication received before hospitalisation (see Table 11. Inequality models of prescriptions after hospitalisation, before hospitalisation and inequality in prescriptions between the two most used alcohol dependence drugs

column 7,8). Receiving disulfiram prior to hospitalisation was associated with an increase in odds of receiving disulfiram after hospitalisation compared to not having prescriptions. Conversely, receiving acamprosate before hospitalisation was associated with a decrease in the odds of getting prescriptions for disulfiram after.

5.6.3 Comparative effectiveness

The comparative effectiveness modelling shows that prescribing disulfiram, compared to acomprosate, was associated with a reduced risk of first alcohol related hospitalisation. All three methods were consistent in their findings (see Figure 8). In both Scot PIS and Eng CPRD, instrumental variable modelling produced point estimates showing larger associations but with wider confidence intervals. Point estimates across the two datasets varied from HR=0.54 (95% CI: 0.33-0.88) for PPP IV in Scot-PIS to HR=0.77 (95% CI: 0.66-0.89) for propensity score in Eng CPRD, signifying that disulfiram was associated with a reduction in the risk of alcohol related hopitalistation between 46.3% and 23.1% comapred to acamprosate. The validation of IV assumption can be seen from Table S7 in supplementary material.



6

Figure 8. Point estimate of hazards ratio for AUD hospitalisation and confidence intervals of the models measuring comparative effectiveness research of disulfiram and acamprosate (using acamprosate as reference).

[6] Acamprosate was the reference variable. Circles are for point estimate related to models on Scot-PIS, squares for Eng-CPRD.

5.6.4 Sensitivity analysis for the unmeasured confounding

It can be seen from Figure S9 in supplementary material for Chapter 5, the potential unmeasured confounder does not need to be strongly associated (for example (0.2, 0.2)) with the treatment to make the estimated treatment effect across the null hypothesis (within the red curve). The model is sensitive to the unmeasured confounding in this case.

5.7 Discussion

We found the prescription rate of alcohol dependence medications was limited to only 2-4% in primary healthcare and 6-7% in secondary healthcare. Our analyses highlighted that several demographic factors were associated with inequalities in prescribing for alcohol dependence. While some factors such as sex and age were associated with inequality similarly after primary and secondary healthcare events, others were drivers for inequality only after a secondary healthcare episode. For example, living in the most socio-economically deprived areas was associated with lower odds of receiving prescriptions within 60 days after the first AUD secondary healthcare episode, but similar results were not observed after a primary healthcare visit. The comparative effectiveness modelling found that disulfiram has reduced risk in preventing the first alcohol related secondary healthcare episode compared to acamprosate. Further, we showed that those living in the least socio-economic deprived areas were associated with an increase in odds of being prescribed the most effective drug (disulfiram) after both primary and secondary healthcare visits. We believe that these findings have important implications for socio-economic health inequalities for the alcohol dependent population.

Our findings are in line with other UK studies, showing a low percentage of pharmacotherapy for patients diagnosed in primary healthcare (Thompson et al., 2017). While the percentage of prescriptions after AUD hospitalisations is generally higher (between 5% and 7% in the overall population), this can still be considered limited as hospitalised patients are likely to have more severe alcohol dependence. It is worth noting that our rates (see Figure 7) are lower than official statistics on alcohol dependence in England (1.37% prevalence in 2018-19 (England, 2021). This is mainly because by using read codes specific for alcohol dependence in CPRD we may have selected only patients with moderate/severe dependence (Thompson et al., 2017) within primary healthcare, which are already a subsample of all the alcohol dependent people estimated in the overall population of England. In addition, under-recording of diagnostic read codes from some practitioners (Tulloch et al., 2020) may have reduced our rates further. While there have been falls in the numbers of people in treatment for alcohol problems in England in recent years (analysis, 2018), they may not fully explain the variation in primary healthcare diagnosis over time we observed (Figure 7). Indeed, it is more likely to be related to the dataset having important variations over the years (new practices added every year, which may have different procedures and attitudes in using Read codes). This hypothesis is supported by the variations in annual presentation rates also found in previous studies utilising CPRD for alcohol dependence analyses (Thompson et al., 2017) and by the contrast with a more linear trend of rates of secondary healthcare AUD episodes built on a nationwide dataset (Figure 7).

Regarding prescription inequality in primary healthcare, Thompson et al. (Thompson et al., 2017) in a similar study utilising CPRD between 1990 and 2013, found comparable inequality patterns for sex and age in determining imbalances in odds in receiving alcohol dependence prescriptions. When we ran the same analysis in secondary healthcare, we found also that socio-economic deprivation status was associated with disparities in receiving prescriptions within 60 days from discharge. However, in contrast, the extent of such disparities decreased for prescribing any time in the future. This could suggest that distinct deprived groups can have different ease and access to care in the initial phase after hospital discharge, which is the most critical period in avoiding relapses (Hunt et al., 1971). Indeed, individuals with alcohol dependence requiring hospitalisation often require specialist alcohol treatment in hospitals or in community settings. Studies describing a lower utilisation of specialist care in groups with lower levels of educational attainment (Stirbu et al., 2011), can explain why we found lower prescription rates in the most deprived areas. We cannot assume that the overall inequality we found in prescriptions concerning sex and age, which are consistent across primary and secondary health care, can be attributed to practitioners or to services prescribing the medications. On the contrary, we believe that a combination of factors such as the lower propensity to seek help of certain patient groups (e.g., males less likely to seek consultation (Wang et al., 2013), especially regarding psychological matters (Liddon et al., 2018) can be responsible for this.

We also found other factors such as comorbidities and previous alcohol dependence medications associated with the odds of receiving prescriptions. Having a history of mental health comorbidities was associated with an increase in the odds of being issued prescriptions before the hospitalisation (see Table S6 in supplementary material) and with a reduction in the odds of getting prescriptions afterwards. This could suggest that patients with certain comorbidities are also more likely to be in contact for mental health assistance and more likely to be treated with alcohol dependence pharmacotherapies aimed to prevent a future hospitalisation. Similarly, our results show that having already received alcohol dependence prescriptions in the past, increased the odds of receiving such prescriptions after hospitalisation. This might indicate that individuals who have already received treatment for alcohol use in the past were more likely to receive it again after hospitalisation compared to those who never had it.

Our analysis of real-world data on a nationwide cohort in Scotland showed that disulfiram is superior to acamprosate in avoiding a first alcohol-related hospitalisation, which was also observed in a representative cohort from England. Our results are in accordance with previous evidence from small randomised control trials (Laaksonen et al., 2008, de Sousa and de Sousa, 2005) and a small observational study (Diehl et al., 2010) that reported disulfiram to be more effective in maintaining abstinence, craving, days until relapse and consumption and abstinence, respectively. Our instrumental variable analysis showing similar results to methods that adjust for measured confounders by indication only, strengthens the internal validity of our study. The wider confidence intervals of the PPP IV models can be ascribed to the nature of IV can only explains a small fraction of the variation in the exposure, so that the standard error of 2SLS will naturally be larger than of OLS (Wooldridge, 2010). The point estimates of the propensity score and covariate adjustment models being closer to the null may be due to a positive correlation between unmeasured confounders (captured by PPP IV) and probability of being prescribed disulfiram, as well as a negative correlation between the unmeasured confounder and outcome. It is worth reminding that we do not link this effect to the pharmacological substances in the drugs only, but it could be generated by a mixture of other factors such as the close monitoring suggested for disulfiram administration. One potential reason is that the severity of alcohol addiction is not properly measured and recorded in the data used in this chapter. Diehl et al. also suggested that the patients who had been prescribed with disulfiram were at more severe situation of alcohol addiction than patients who had been prescribed with acamprosate. Patients with higher expected health benefit (generally healthier) tend to receive acamprosate (Diehl et al., 2010).

In our final inequality analysis (see Table 11. Inequality models of prescriptions after hospitalisation, before hospitalisation and inequality in prescriptions between the two most used alcohol dependence drugs

, models 3, 4 and 7, 8), we showed how living in the most deprived areas decreased the odds of being prescribed the most effective medication to avoid alcohol related hospitalisation compared to living in the least deprived areas. This remained the only driver of prescription imbalances between the two drugs, and it was consistent between primary and secondary healthcare. We believe this has important implications for health inequality. Again, we do not attribute this to hypothetical prescribers' bias, but more likely due to unmeasured factors such as less available assistance, supervision or close clinical monitoring (recommended for disulfiram (Excellence)) in individuals living in more deprived areas. The general inequality of prescriptions for alcohol dependence combined with the inequality of the most effective medications in favour of the least deprived groups can partially explain the social imbalance of the burden of alcohol. To address this, we believe that improving patient access to specialist services after being hospitalised for alcohol related reasons and developing new integrated care pathways are essential especially for hard-to-reach patients.

5.8 Strengths and Limitations

We found new findings regarding inequality especially related to the prescriptions between acamprosate and disulfiram, with relevance for care and support plans of alcohol dependent patients in both primary and secondary healthcare. In our opinion, we also provided the best comparative effectiveness evidence to date based on real-world data. We utilised two datasets referring to nationwide (Scot-PIS) and highly representative (Eng-CPRD) populations. Results between datasets and across methods were consistent. In essence, this triangulation internally cross-validated our comparative effectiveness study. Previous real-world studies had lower power (Diehl et al., 2010) and were not supported by any such triangulation strategy. Related to this, when we built in two different datasets an analogous cohort (for the comparative effectiveness analysis) our results were consistent. In contrast, when we built different cohorts between datasets (for the inequality analysis), we found diverse results that we interpreted as differences between primary and secondary healthcare. We believe that this accordance between cohorts and results across datasets corroborates the reliability of our results and interpretations.

Whilst both Scot-PIS and Eng-CPRD are two datasets referring to the same country (UK), they reflect two different nations which, even if similar, identify different populations. Having both datasets on primary and secondary healthcare referring to the same population would have been ideal, but unfortunately primary healthcare data are not available in Scotland, and we did not have access to nationwide secondary healthcare data for England. Despite the minor differences across populations in general and in particular regarding individuals with AUD (Office for national statistics 2022) , we believe that our triangulation offering similar results for same cohorts between datasets as well as multiple models for both primary and secondary healthcare confer robustness to our results and interpretations. Moreover, additional practices added every year into the Eng-CPRD dataset did not allow a full look-back period for every patient, guaranteeing the certainty of measuring 'the first diagnosis for alcohol dependence in primary healthcare' only for a restricted number of patients. However, stable outcomes over the years such as the percentage of patients receiving prescriptions despite variations in the number of practices introduced in the dataset should not make this a major source of concern. Another potential limitation was that for Scot-PIS we looked at all the ICD-10 codes identifying AUD hospitalisations rather than limiting our analysis only to alcohol dependence like we did for Eng-CPRD where we utilised read codes for alcohol dependence only. We included all AUD diagnoses in Scot-PIS mainly to correct for possible errors in recording data across different alcohol related diagnostic codes which are possible in general/acute hospital records. Indeed, in Scot-PIS some of the people not hospitalised for alcohol dependence but for other AUD conditions (e.g., withdrawal or intoxication) received alcohol dependence prescriptions, while this did not happen in English primary healthcare for patients without an alcohol dependence diagnosis. We are also aware that some potentially key variables were not always considered across our analyses (e.g., comorbidities in primary care data), but we did not have access to read codes for conditions not related to alcohol. Finally, as already mentioned, utilising alcohol dependence read codes may have selected our population to more severe patients, with a potential minor generalisation of our conclusions to mild alcohol dependence individuals.

5.9 Conclusion

Alcohol dependence medications are not extensively prescribed in the UK. Inequalities in prescribing alcohol dependence medications exist, especially across sex, age and socio-economic deprivation groups. The extent of such inequality is different in primary and secondary health care settings. Disulfiram is superior to acamprosate in avoiding alcohol related hospitalisations in large, unselected observational datasets. Further, there is inequality within alcohol dependence prescriptions, with individuals living in the most deprived areas having lower odds of being prescribed the most effective drug. This has implications for health inequality highlighting the need of building new strategies to reduce the societal imbalance in the burden of alcohol.

Chapter 6. Comparing DPP-4 inhibitor and sulfonylurea as second-line treatment for T2DM patients: a target trial emulation

6.1 Publication details

This article has not been submitted for publication.

6.2 Data used in this chapter.

The concept of this chapter is built from the NIHR-funded PERMIT study (see PERMIT | LSHTM on https://www.lshtm.ac.uk/research/centres-projects-groups/permit) which aims to assess the long-term effectiveness and budget impact of alternative second-line drug treatments for patients with type 2 Diabetes Mellitus. The application form for the access of Scottish Diabetes Research Network (SDRN) is attached in supplementary material for Chapter 6. Data tables and variables used in this chapter are listed below.

Data tables	Variable	Variable Description
s_prescription	serialno	Unique identifiers of the patients
	drugname_clean	Name of drugs
	pseudonymised_populationid	Pseudonymized data source identifier refers to the location
	startdate	Prescription date
	concept_id	Concept id for matching with UID of o_concept_drugs
	bnfcode	BNF code of the drugs (BNF codes used in this chapter is 6.1 :)
o_concept_drugs	UID	The unique record identifier for this table
	bnfcode	The British National Formulary for the product selected by GP
	drugname	Name of the drug
	strength	The strength field provides the dosage of the drug per item
o_person	serialno	Unique patient identifier
	gender	Patient's gender
	date_of_birth	Date the birth
	ethnic	Patient's ethnicity
	serialno	Unique patient identif
o_observation	concept_id	This is the id number of the concept in o_concept_observation that describes the observation being recored.
	num_value	Observation values (e.g., HbA1c level)
O_concept_observation	date	Date of the records
	UID	Identifier for this observation concept to be matched with concept_id in the o_observation table. HbA1c: 1001,8002 BMI: 1123 Derived eGFR: (4012,4013,4014,4015,4016,4017,4020,4021)

6.3 Abstract

Background

There are variations in the choice of second-line treatment, namely DPP-4 inhibitor or SU, after metformin monotherapy fails to reduce the HbA1c level for type 2 diabetes patients. This chapter aims to provide real-world evidence on the comparative effectiveness of SU and DPP-4 inhibitor in reducing HbA1c level for type 2 diabetes patients.

Method

This study is a target trial emulation where the cohort is designed based on the inclusion and exclusion criteria from existing RCTs. The target trial emulation study was conducted using a nationwide dataset, Scottish Diabetes Research Network (SDRN), with a study period between 2014 and 2019. The multivariable regression approaches as well as IV approach is conducted to estimate the short-term effectiveness of DPP-4 inhibitor and SU in reducing the HbA1c level. Outcomes include reduction of HbA1c level (numerical outcome), and whether the HbA1c level is reduced to less than 42 mmol/mol or between 42 mmol/mol to 47 mmol/mol (binary outcomes).

Result

The mean difference from multivariable regression models indicates that SU is associated with 2.56 units more reduction in HbA1c levels than DPP-4 inhibitor (2.56, 95% CI: 1.07-4.04). For the IV approach, the result from 2SLS is not statistically significant. SU is associated with higher probability of reducing HbA1c level to less than 42 mmol/mol in multivariable logistic regression (OR: 1.98, 95% CI:1.50-3.73) and 2SRI approach (OR: 8.17, 95% CI:1.02-68.48).

Conclusion

I found real-world evidence supporting SU as a superior treatment to DPP-4 inhibitor in reducing HbA1c levels, which supports that SU should remain in the treatment portfolio as the second-line treatment for T2DM patients. PPP performs as a valid IV and give statistically significant results for a binary representation of the outcome but not for a numerical representation.

Keywords: target trial emulation, instrumental variable, prescribing preference, type2 diabetes, DPP-4 inhibitor, sulfonylurea, second-line treatment

6.4 Introduction

In decision-making in health service research, RCT is considered as a reliable source of evidence (Greenland, 1990). However, RCTs are not always feasible or ethical. Observational comparative effectiveness studies which use real-world data are an alternative to RCTs. However, such studies need to overcome unmeasured confounding (confounding by indication) otherwise they will have poor internal validity. Many analytical methods can be used, including instrumental variable (IV) approaches which directly account for unmeasured confounding. However, investigators of observational studies tend to pay more attention on analysis, rather than design, including eligibility criteria, start of follow-up, etc. This may cause severe bias in the estimation of treatment effect (Gomes et al., 2022). On grounds of that, target trial emulation has been proposed to mimic RCT eligibility criteria using real-world data which can improve the design of non-randomised studies (Hernán and Robins, 2016). A target trial emulation applies the principles from RCT on observational data so that it can emulate the RCT in an observational data setting, usually with larger sample sizes (Labrecque and Swanson, 2017).

According to NICE guidelines, metformin monotherapy is usually the first-line treatment for the type 2 diabetes mellitus (T2DM) patients. If the HbA1c level is inadequately controlled by metformin monotherapy, second-line treatments can be added after it. Commonly prescribed second-line antidiabetics drugs are dipeptidyl peptidase-4 inhibitor (DPP-4 inhibitor) and sulfonylureas (SU), and sodium/glucose cotransporter 2 (SGLT2).

Recent meta-analyses of RCTs suggest that DPP-4 inhibitor is not statistically different with SU in terms of clinical efficacy. While DPP-4 inhibitor users are less risky in developing hypoglycaemic events (Esposito et al., 2011, Foroutan et al., 2016). However, another meta-analysis of RCT conclude that the DPP-4 performs better in maintaining the durability of controlling HbA1c level than SU (Chen et al., 2018).

In terms of the effectiveness, recent observational studies using large scale electronic health data (Tan et al., 2021, Lee et al., 2022) that compared the effectiveness and safety of second-line treatments for T2DM. Besides, observational study evidence based target trial emulation approach has been published (Bidulka et al., 2021). This study compared the effectiveness of DPP-4 inhibitor, SU and SGLT-2 inhibitor. It also conducted IV methods to attempt to eliminate the potential unmeasured confounding. However, this study utilised the data from clinical practice research datalink (CPRD) which although is deemed a representative routinely-collected dataset it is not nationwide, so selection bias is a concern. Despite the fact that the latest NICE guideline recommends SGLT-2 inhibitor for patients with higher risk of cardiovascular diseases (CVD) (NICE guideline 2022) , SGLT-2 inhibitor prescriptions are less than DPP-4 inhibitor and SU in routine health data set. Therefore, this chapter is conducted from a pragmatic perspective and focused on the comparison between SU and DPP-4 inhibitor in a nationwide cohort.

The primary objective of this chapter is to implement the IV method in observational comparative effectiveness research based on a nationwide health dataset from Scotland. The secondary objective is to add observational evidence regarding the effectiveness of DPP-4 inhibitors and SU. It is arranged by two approaches:

1. In a target trial emulation, compare the short-term effectiveness of two second-line treatments, DPP-4 inhibitors and SU, as second-line treatment after metformin monotherapy for T2DM patients.

2. Using the data defined from scenario 1, compare the short-term effectiveness of DPP-4 inhibitors and SU by defining binary outcomes using whether or not the treatment reduces the HbA1c level.

6.5 Approach 1

6.5.1 Method

The data is from a Scottish national population-based register (https://www.sci-diabetes.scot.nhs.uk/), called the Scottish Diabetes Research Network (SDRN). This is a dynamic clinical information system which contains detailed clinical records, including BMI, eGFR, HbA1c, for all the patients in Scotland who have been diagnosed with diabetes (McGurnaghan et al., 2022). This target trial emulation study is designed following the guideline provided by Hernán and Robins in 2016 (Hernán and Robins, 2016). This protocol is based on the Nauck et al's research (Nauck et al., 2007) which is a double-blind, non-inferiority randomised clinical trial. I replicated their selection strategy but made some adjustments due to availability of data in SDRN (see Table 12).

Inclusion criteria	• Adults aged from 18 to 78, with type 2 diabetes mellitus on metformin monotherapy. • First prescription of second-line treatment between 1 January 2014 to 31st December 2021. • First prescription of second-line treatment is within 60 days after metformin monotherapy. • Earliest date mentioned is before 31 Dec 2018 (make sure follow-up time is longer than 1 year).
Exclusion criteria	• History of type 1 diabetes • eGFR <30 • pregnancy within 12 months of second-line treatment initiation • Prescribed with insulin within 8 weeks
Treatment strategies	• SU includes glibenclamide, gliclazide, glimepiride, glipizide, tolbutamide. • DPP-4 inhibitor includes linagliptin, saxagliptin, sitagliptin, vildagliptin.
Assignment procedure	• Prescription of the second-line treatment on SDRN data. • Accounting for confounding: Adjusted OLS; propensity score, two-stage least square using last year prescribing preference of hospital or GP practice etc. as instrumental variable.
Day zero	• Date of the first prescription of DPP-4 inhibitor or SU
Follow-up period	• Start at the date of the first prescription of DPP-4 inhibitor or SU and ends at 365 days after the first prescription. With 3-month time-window on both sides. We selected the HbA1C records which is closest to the prescribing date as the baseline value if there are multiple records for one patient.
Outcome	• Change of HbA1c 1 year after the first prescription. Time-window is 3 months (from 40 weeks to 64 weeks).
Causal contrast of interest	• Intention-to-treatment effect. • Ignore switching

Table 12. Protocol for target emulation trial

The flow chart of cohort definition is shown in Figure 9. As can be seen, there is approximately a 60 % reduction from the original cohort defined by the T2DM patients who had second-line treatment as adds-on for the metformin from 1st January 2014 to 31st December 2021 (N=9532). The total number of the study population is 3789.

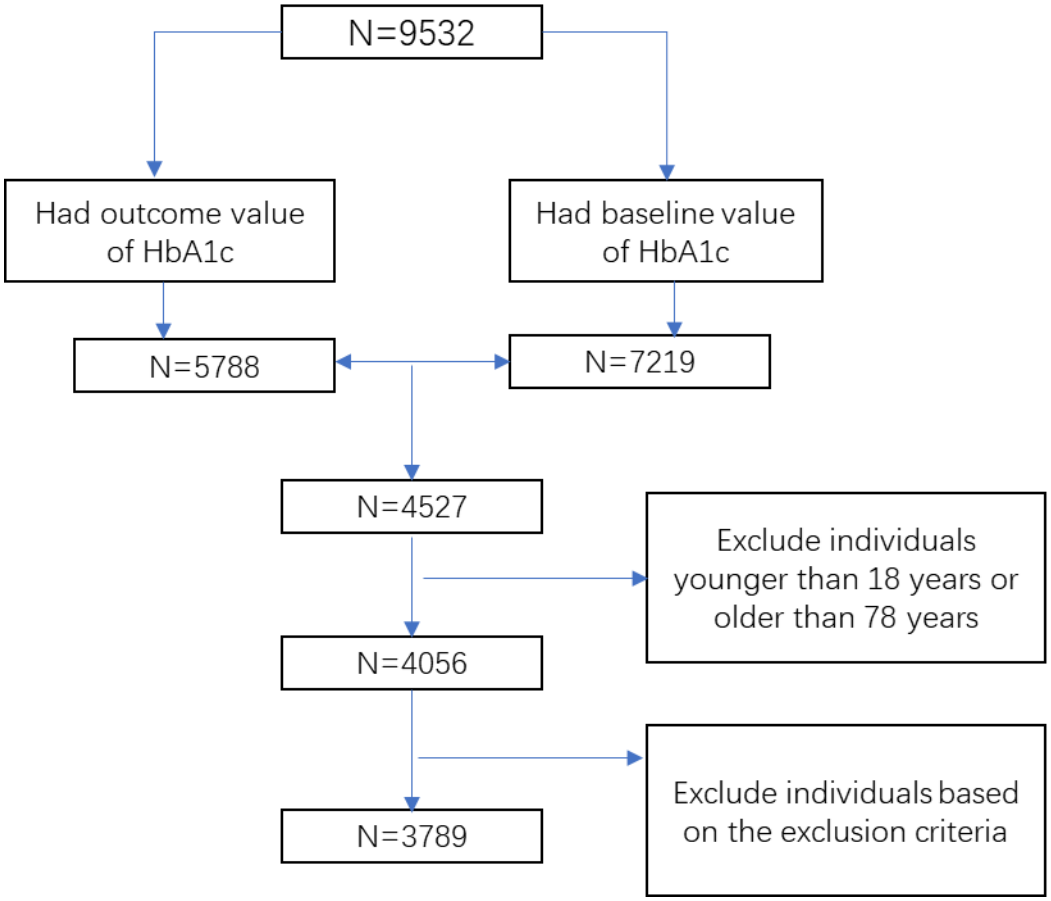


Figure 9. Flow chart of defining study cohort

I conducted OLS, 2SLS and propensity score matching, using inverse propensity score weighting (IPSW) as conventional multivariable adjusted models. In term of IV method, 2SLS is the most commonly used estimation method for IVs and is well suited for numerical outcome variables (such as change in HbA1c). As I mentioned in Chapter 2, 2SLS can only estimate the local average treatment effect (LATE) when the monotonicity assumption hold. For that reason, I also estimated marginal treatment effect (MTE) for the whole population. MTE considers the probability of receiving certain treatment (D=1) given a value of numerical instrumental variable (Z) and covariates (X) the as the propensity score: P(z)=Pr (D=1|Z=z, X=x). Then the outcome given the treatment received can be depicted as a function of X and P(z): E (Y|X=x, P(Z)=p). MTE can be obtained nonparametrically using local instrumental variable estimator by calculating the partial derivate of E (Y|X=x, P(z)=p) with respect to the propensity score.

In SDRN, the identifiers for physicians are not recorded. Therefore, the proxy for the PPP cannot be identified. Alternatively, I exploited one-year prescription history at the level of general practice or hospital as the proxy of prescribing preference. The proxy was calculated as the proportion of SU prescriptions among one-year prescriptions prior to the first prescription date of second-line treatment.

Covariates

Covariates include the baseline HbA1c level, age at the first prescription of second-line treatment, gender, body mass index (BMI) before the first prescription of second-line treatment and comorbidity (in the form of CCI and in quintiles). The hospitalisation record was obtained from SMR01. The reason for the hospitalisation was recorded in the form of ICD-10. I conducted multiple imputation (MI) for the missing values in BMI which was implemented by classification and regression trees with 20 imputations (Tierney et al., 2015), conducted by ‘mice’ package in R (Van Buuren and Groothuis-Oudshoorn, 2011).

Validation of assumptions

The PPP IV was dichotomised at the median (0.75). The relevance assumption was validated using the F-statistics. The monotonicity assumption was verified using the monotonicity inequality which were proposed by Balke and Pearl (Balke and Pearl, 1997) . The independence assumption was validated using covariate balance presented by standardised mean difference (SMD). Variables with the SMD less than 0.1 are considered as balanced variables (Austin, 2009).

6.5.2 Results

6.5.2.1 Validation of IV assumptions

The F-statistics of the instrumental variable is 101.765 (see Table), far exceeding the 10 threshold and indicating a strong instrumental variable (Stock and Yogo, 2002). Figure S10 in supplementary material indicates that the dichotomised PPP IV reduced the SMD of all covariates to under 0.1 which can be treated as the IV reduces the imbalance among covariates indicating the independence assumption is likely to hold.

6.5.2.2 Descriptive statistics of the study population.

After applying the inclusion and exclusion criteria in the protocol, there are 3789 individuals in Scotland who have been diagnosed with T2DM and received second-line treatment of DPP-4 inhibitor or SU between 2014 and 2019. SU was much more prescribed that DPP-4 inhibitor (four times more). Patients treated with SU had on average, higher levels of baseline HbA1c than patients treated with DPP-4 inhibitor. This may be because SU is a more traditional choice than DPP-4 inhibitor. Besides, SU group is more comorbid than the DPP-4 inhibitors group which is consistent with the baseline characteristics found in other studies (Chung et al., 2019, Eriksson et al., 2016). There are 2.7% of missing values in BMI (see Table 13).

	dpp4i (N=718)	su (N=3071)	Overall (N=3789)
Age (years)			
Mean(SD)	59.9(10.6)	60.1(11.4)	60.0(11.2)
Gender			
Female	267 (37.2%)	1144 (37.3%)	1411 (37.2%)
Male	451 (62.8%)	1927 (62.7%)	2378 (62.8%)
Charlson Comorbidity Index			
Mean(SD)	0.260(0.542)	0.364(0.668)	0.345(0.647)
Ethnic			
Asian	43 (6.0%)	194 (6.3%)	237 (6.3%)
Black	7 (1.0%)	25 (0.8%)	32 (0.8%)
Missing/Unknown	100 (13.9%)	473 (15.4%)	573 (15.1%)
Mixed/Other	8 (1.1%)	55 (1.8%)	63 (1.7%)
White	560 (78.0%)	2324 (75.7%)	2884 (76.1%)
Baseline HbA1c (mmol/mol)			
Mean(SD)	70.5(20.0)	78.4(26.1)	76.9(25.2)
BMI(kg/m2)			
Mean(SD)	33.6(7.02)	32.8(6.85)	32.9(6.89)
Missing	11 (1.5%)	93 (3.0%)	104 (2.7%)

Table 13. Descriptive statistics of the study population

6.5.2.3 OLS and 2SLS estimates

Although I followed the current observational studies to decide the covariates used in the models, Tennant argued that adjusting the baseline value is source of biased causal inference, except for in the RCT or the baseline value is not mediating the exposure, or being the competing risks (Tennant et al., 2021). While Glymour argued that this conclusion cannot be generalised (Glymour, 2022). Therefore, I presented results with or without adjusting baseline HbA1c in Table 14.

Comparison between SU and DPP4i

	OLS_1		OLS_2		Propensity score 1		Propensity score 2		2SLS 1		2SLS 2	
Predictors	*Estimates*	*CI*	*Estimates*	*CI*	*Estimates*	*CI*	*Estimates*	*CI*	*Estimates*	*CI*	*Estimates*	*CI*
Drug [su]	2.56 ***	1.07 – 4.04	8.89 ***	6.82 – 10.96	3.13 ***	1.40 – 4.86	8.91 ***	6.80 – 11.02	-4.78	-15.96 – 6.39	-1.59	-19.12 – 15.94
gender [Male]	1.30 *	0.01 – 2.59	-0.91	-2.76 – 0.94	0.88	-0.83 – 2.58	-1.26	-3.38 – 0.86	1.29	-0.01 – 2.59	-0.98	-2.84 – 0.89
age	0.16 ***	0.10 – 0.23	-0.35 ***	-0.44 – -0.26	0.18 ***	0.09 – 0.27	-0.26 ***	-0.36 – -0.16	0.17 ***	0.10 – 0.23	-0.35 ***	-0.44 – -0.26
ethnic [Black]	8.31 *	1.61 – 15.02	11.82 *	1.03 – 22.61	2.72	-8.61 – 14.06	8.73	-0.06 – 17.51	8.00 *	1.08 – 14.91	11.45 *	0.29 – 22.60
ethnic [Missing/Unknown]	4.81 **	1.89 – 7.72	13.94 ***	10.13 – 17.75	3.69 *	0.12 – 7.26	11.88 ***	7.52 – 16.24	4.77 **	1.84 – 7.70	14.11 ***	10.29 – 17.94
ethnic [Mixed/Other]	1.40	-3.30 – 6.10	5.21	-1.82 – 12.24	-2.40	-8.29 – 3.49	1.10	-5.35 – 7.55	1.77	-3.10 – 6.64	5.84	-1.41 – 13.09
ethnic [White]	2.18	-0.40 – 4.75	6.60 ***	3.37 – 9.84	0.70	-2.39 – 3.79	5.59 **	1.87 – 9.30	2.07	-0.51 – 4.65	6.56 ***	3.34 – 9.78
bmi	-0.07	-0.18 – 0.03	-0.04	-0.19 – 0.11	-0.09	-0.22 – 0.04	-0.05	-0.21 – 0.11	-0.09	-0.20 – 0.02	-0.07	-0.22 – 0.09
cci [1]	0.69	-0.78 – 2.16	5.17 ***	2.90 – 7.43	-0.22	-2.31 – 1.86	3.63 **	1.00 – 6.26	0.81	-0.72 – 2.33	5.45 ***	3.07 – 7.83
cci [2]	0.36	-2.78 – 3.49	6.74 **	1.93 – 11.56	2.41	-3.11 – 7.92	8.30 **	2.06 – 14.54	1.11	-2.11 – 4.34	8.00 **	2.83 – 13.17
cci [3]	0.93	-4.23 – 6.08	1.57	-5.74 – 8.88	4.34	-1.05 – 9.74	3.53	-3.95 – 11.01	1.23	-3.88 – 6.34	2.03	-5.25 – 9.30
cci [4]	2.41	-8.90 – 13.71	2.57	-18.32 – 23.47	1.93	-9.57 – 13.44	1.32	-19.82 – 22.45	3.91	-7.51 – 15.34	4.76	-16.27 – 25.79
cci [5]	-37.68 ***	-40.01 – -35.36	-22.21 ***	-34.03 – -10.39	-37.79 ***	-41.08 – -34.51	-23.10 **	-39.43 – -6.78	-36.34 ***	-39.40 – -33.28	-19.87 **	-32.45 – -7.30
baseline hba1c	0.84 ***	0.81 – 0.87			0.81 ***	0.76 – 0.86			0.85 ***	0.81 – 0.89		
Observations	3789		3789		3789		3789		3789		3789	
R^2	0.566		0.053		0.564		0.050		0.556 / 0.554		0.031 / 0.028	

** p<0.05 ** p<0.01 *** p<0.001*

[7]

Table 14. OLS, propensity score and 2SLS estimates.

In Table 14, the OLS inferential results show that those prescribed SU have, on average, more HbA1c reduction compared to those prescribed DPP-4 inhibitors (2.56, 95% CI: 1.07-4.04). The baseline HbA1c value is highly associated with the change of HbA1c. Patients with one unit higher in baseline HbA1c level is more likely to have 0.84 more unit in the reduction of HbA1c (95% CI: 0.81-0.87). In the case of not adjusting the baseline value, the treatment effect from SU is statistically significant with a level of HbA1c 8.89 units higher on average in the reduction of HbA1c than DPP-4 inhibitors. This trend is also seen in the propensity score results. 2SLS show the opposite results indicating DPP-4 inhibitors is associated with more reduction in HbA1c, but with wide confidence intervals. Note that, for the estimates from 2SLS, adjusting or not adjusting baseline HbA1c value does not show strong evidence of difference, as they are statistically insignificant with wide confidence intervals.

[7] OLS_1 and 2SLS_1 are the models adjusting baseline HbA1c. While OLS_2 and 2SLS_2 are models without adjusting baseline HbA1c.

6.5.2.4 Marginal treatment effect

The marginal treatment effect results are shown in Table 15. Note that 2SLS estimates the LATE for the complier group where the treatments received are determined by the instruments. I conducted MTE methods to extrapolate LATE nonparametrically using the same dichotomised PP IV. The IV estimates show that the patients who were given SU had less reduction in HbA1c. Neither ATE nor LATE is statistically significant. The difference between LATE and ATE is due to heterogeneity of treatment effects that affects the treatment decision (Fang et al., 2012). For example, patients with higher expected benefits tend to get the medication. It can be seen from Table 15 after adjusting for baseline HbA1c, the estimated LATE became larger than the estimated ATE. One possible reason is that the decision of the treatment changed after taking the baseline HbA1c level into account. The noticeable difference between the LATE without adjusting baseline HbA1c and LATE adjusting baseline HbA1c may suggest violations of IV assumptions. Note that the estimates from LATE and ATE are not statistically significant, there is no strong evidence of differences between these two approaches.

Model	LATE (95% CI)	ATE (95% CI)
Without adjusting baseline HbA1c	-5.862 (-28.913- 18.773) p value: 0.649	-16.880 (-55.009-23.107) p value :0.353
Adjusting baseline HbA1c	-26.207 (-56.514- 3.197) p value: 0.094	-16.205 (-38.241-4.050) p value: 0.144

Table 15. Average treatment effect and local treatment effect estimated by MTE methods. 95% CI was bootstrapped by 1000 times.

6.5.2.5 Sensitivity analysis for the unmeasured confounding

A sensitivity analysis for the unmeasured confounding was conducted. The sensitivity analysis was conducted using R function ‘sensemakr’(Cinelli et al., 2021). As it shows in Figure 11 in supplementary material, the potential unmeasured confounder needs to explain at least 5.2% of the residual variance of treatment and outcome to make the estimated treatment cross the null hypothesis. The result from Wu-Hausman test is presented in Table S9. It shows that null hypothesis cannot be rejected. There is no significant difference between OLS estimate and 2SLS estimate.

6.6 Approach 2

6.6.1 Method

Two binary outcomes are defined to measure the effectiveness of SU and DPP-4 inhibitors in reducing HbA1c level (Canivell et al., 2019). One binary outcome is defined based on whether the HbA1c level had been reduced to less than 42mmol/mol at which point non-pregnant patients can be considered free of T2DM. Another binary outcome is defined as whether HbA1c level had been reduced to less than 47 mmol/mol which is considered as at high risk of T2DM or prediabetes (NICE Guideline 2022). The statistical methods are logistic regression and 2SRI. As one of the nonlinear extensions of the two-stage approaches, 2SRI is used to estimate the treatment effect for the binary outcomes. The first stage of 2SRI is the same as in the 2SLS, while the second stage adds the residual from the first stage as an additional covariate instead of using the predicted treatment from the first stage. Because of the noncollapsibility of OR, the stratified ORs may not be the weighted average sum of unadjusted OR (Pang et al., 2016), the results from unadjusted logistic regression are also presented (Schuster et al., 2021).

6.6.2 Results

6.6.2.1 Validation of IV assumptions

In terms of the binary outcomes, I conducted the validation of monotonicity assumption and independence assumption using the monotonicity inequality and IV inequality described in section 2.2.4. As it shows in Table S9 in supplementary material, using 1-year prescription history as IV may violate the monotonicity assumption when HbA1c less than 42 mmol/mol as outcome. I conducted the same test for different forms of IV and found out that using prior 200 days prescription as IV make both IV equality and monotonicity inequality hold.

6.6.2.1 non-IV results from logistic regression

As it can be seen from Table 16, SU is associated with higher odds compared to DPP-4 inhibitor of being reduced to less than 42 mmol/mol (OR= 1.98). The unadjusted logistic regression shows similar estimated OR. Note that SU is not shown effective in reducing the HbA1c to prediabetic state.

Outcome	Unadjusted logistic regression (95% CI)	Adjusted logistic regression. (95% CI)
HbA1c level less than 42 mmol/mol	2.42 (1.42-4.53) p value： 0.003	1.98 (1.15-3.73) p value: 0.022
HbA1c level between 42 to 47 mmol/mol.	1.38 (1.05-1.86) p value: 0.027	1.22 (0.91-1.65) p value: 0.188

Table 16. Estimated OR from logistic regression.

6.6.2.2 2SRI results

As I previously mentioned, given that 1-year prescription may not be valid IV in the validation of monotonicity assumption, I conducted 2SRI for multiple forms of IV by taking different lengths of prescription history into account. It can be seen from Table 17, the estimated OR using prior 200 days as PPP IV is not much different from that using 1-year prescription as IV. 2SRI estimates show that SU is associated with higher odds of reducing HbA1c to less than 42 mmol/mol. At the same time, the SU is associated with lower odds of reducing HbA1c to 42 to 47 mmol/mol.

The proportion of SU prescribed during N days before the first prescription of second-line treatment (IV)	Strength of IVs (in F-statistics)	2SRI estimate (OR) for reducing HbA1c level less than 42 mmol/mol. (95% CI)	2SRI estimate (OR) for reducing HbA1c level to between 42 to 47 mmol/mol. (95% CI)
N=120	52.964	2.943 (0.215-40.808) p value: 0.417	0.333 (0.0785-1.372) p value: 0.131
N=200	74.193	5.289 (0.533-45.540) p value: 0.136	0.245 (0.0581-0.967) p value: 0.0499
N=365	101.765	5.097 (0.570-45.679) p value:0.142	0.288 (0.0800-0.994) p value: 0.0524
N=480	92.817	6.189 (0.787-45.610) p value: 0.0743	0.223 (0.0597-0.789) p value: 0.0228
N=600	110.970	8.185 (1.023- 68.478) p value: 0.0476	0.190 (0.0534-0.648) p value: 0.00906

Table 17. 2SRI estimates of binary outcomes.

6.7 Discussion

The results from 2SRI and logistic regression provide real-world evidence that there is a difference in effectiveness of SU and DPP-4 inhibitors in reducing HbA1c level as second-line treatments for T2DM in Scotland from 2014 to 2021. It is inconsistent with the conclusion from the benchmark RCT used in this chapter (RD: 0.20, 95% CI: -0.90-1.30) which indicate that there is no statistical difference in the effectiveness of SU and DPP-4 inhibitor (Nauck et al., 2007). There are also inconsistencies in the wider evidence base where DPP-4 inhibitor is proven not superior (Esposito et al., 2011, Foroutan et al., 2016). Besides, there are also studies suggesting that DPP-4 inhibitor is superior to SU in controlling the glycaemic level (Fadini et al., 2018). DPP-4 inhibitor may be more cost-effective than SU (Ruan et al., 2022) and decreases more BMI (Gottlieb et al., 2017).

In some studies, SU is considered riskier in terms of causing hypoglycaemia (Khunti et al., 2021, Foroutan et al., 2016). However, there are researchers that hold the opposite view (Mohan et al., 2020). Despite SU being shown to be associated with higher risk of CVD (Khunti et al., 2021, Wang et al., 2022), a more recent cohort study from Scottish National data proved that SU is not associated with higher risk of cardiovascular death or all-cause death than the DPP 4-inhibitor and SGLT2-inhibitor (Wang et al., 2023). They tend to agree SU should remain inside the treatment portfolio. Besides, results from a nationwide cohort from Korea indicated the DPP-4 inhibitor does not show statistical difference in the risk CVD death and renal outcome and associated with higher risk of hospitalisation of heart failure in comparison with SU (Kim et al., 2019).

In terms of the difference in effect sizes in Nauck's study and this chapter, I critically appraise it from two different aspects. First, the study cohort is not totally comparable with Nauck's study. I included Charlson Comorbidity Index (CCI) in the dataset to adjust for more potential confounders (Hernán et al., 2016). Additionally, I included four types of DPP-4 inhibitor and five types of SU which are available in SDRN to enlarge the cohort sample size (listed in the Table 12). I did not consider whether the patients switched back to metformin monotherapy after second-line treatment within a short period of time for the same purpose (i.e., adhered to intention to treat principle). I also relaxed several criteria to increase sample sizes. I set a three-month time-window for the one-year follow-up HbA1c. Secondly, there are essential differences between RCT and target trial emulation. I cannot emulate 'double blindness' of RCTs in observational studies (Labrecque and Swanson, 2017). Another important difference between RCT and target trial emulation is that eligibility assessment and treatment assignment happen at the same time (at time zero) in RCT but not in target trial emulation. When interpreting the discrepancy between the results from RCT and target trial emulation, one need to consider the inconsistency between time zero and time of eligibility assessment which is a cause of bias, including immortal time bias (Hernán et al., 2016).

In terms of whether or not to adjust for the baseline HbA1c as a covariate, I presented results in both ways and the results were sensitive to this choice. Since the baseline HbA1c level is highly associated with the change of follow-up HbA1c (Scheen, 2020), I tend to agree with the results with the baseline HbA1c need to be adjusted for in the regression model.

In this chapter, I conducted 2SLS and 2SRI for the numerical outcome and binary outcome respectively. There are discussions on the comparison between 2SLS and 2SRI in multiple settings where 2SLS is widely proven consistent in the estimation of LATE but tend to be inconsistent in the estimation of ATE (Chapman and Brooks, 2016, Basu et al., 2018, Terza, 2018). They pointed out non-linear 2SRI are more likely to be biased in terms of estimating LATE (Chapman and Brooks, 2016). However, this can be improved by choosing the right forms of residuals (Terza, 2018, Garrido et al., 2012). It shows a great difference in terms of estimation of treatment effect between 2SLS and 2SRI. The estimation from 2SRI is statistically significant and comparable with results from logistic regression. 2SLS does not show any effectiveness differences while 2SRI reveals noticeable difference. This inconsistency indicates that investigators may need to consider different forms of outcome when implementing IV approaches. However, they also need to take the limitations of non-linear 2SRI into account, for example the functional assumption which is hard to validate in terms of using binary IVs (Terza, 2018).

6.8 Strengths and Limitations

One strength of this study is that SDRN is a national data including all patients with T2DM in Scotland, therefore no selection bias exists due to, for example, location or socioeconomic status. Another strength is that this study implemented target trial emulation to make the study cohort closer to RCT. Besides, I considered a stronger list of covariates by adding baseline HbA1c and CCI. I also considered different forms of the outcomes which gave both relative risk and absolute risk difference representations.

One limitation of this study is that I only investigated two types of second-line treatments. A relatively new type of second-line treatment, SGLT-2 inhibitor, has been shown to be more effective than DPP-4 inhibits and SU in some studies (Wilding et al., 2018). In the routine data set used in this study (follow-up from 2014 to 2019), SGLT-2 inhibitor had not been prescribed much in Scotland until 2021. A further limitation is that I focused on the intention-to-treat estimator without considering the cases of switching treatments. According to a recent cohort study in the US, the discontinuation rate for SU is high while DPP-4 inhibitors have better persistence rate (Tan et al., 2021). Apart from the consideration of the strength of the CER, there are studies indicate that the reduction of HbA1c level should not be a measurement of clinical efficacy of add-ons to metformin (Scheen, 2020).

6.9 Conclusion

This target trial emulation concludes that SU is statistically significant superior to DPP-4 inhibitors in controlling HbA1c level after adjustment for measured potential confounding. SU has been shown to be more effective in reducing HbA1c level to less than 42 mmol/mol level than DPP-4 inhibitors in logistic regression and 2SRI for the binary outcome. The faculty-level prescribing preference can be a valid instrument, however, 2SLS and MTE did not give statistically significant results due to the sample size not being large enough to accommodate the IV approach.

Chapter 7. Assessing the performance of Physician's Prescribing Preference as an instrumental variable in Comparative Effectiveness Research with moderate and small sample sizes: a simulation study

7.1 Publication details

- This chapter had been submitted for publication in Journal of comparative effectiveness (under review)
- Abstract of this work has been presented at the 13th Asian Conference on Pharmacoepidemiology (ACPE 2021). Record can be found on the website: https://www.asianpharmacoepi.org/wp-content/uploads/2022/12/ACPE13-ePoster-abstracts.pdf.

7.2 Abstract

Background

Instrumental variable (IV) analyses are used to account for unmeasured confounding in Comparative Effectiveness Research (CER) in pharmacoepidemiology. To date, simulation studies assessing the performance of IV analyses have been based on large samples. However, in many settings, sample sizes are not large.

Objective

In this simulation study, we assess the utility of PPP as an IV for moderate and smaller sample sizes.

Methods

We designed a simulation study in a CER setting with moderate (around 2500) and small (around 600) sample sizes. The outcome and treatment variables were binary, and three variables were used to represent confounding (a binary and a continuous variable representing measured confounding, and a further continuous variable representing unmeasured confounding). We compare the performance of IV and non-IV approaches using two-stage least squares (2SLS) and ordinary least squares (OLS) methods, respectively. Further, we test the performance of different forms of proxies for PPP as an IV.

Results

The PPP IV approach results in a percent bias of approximately 20%, while the percent bias of OLS is close to 60%. The sample size is not associated with the level of bias for the PPP IV approach. However, smaller sample sizes led to lower statistical power for the PPP IV. Using proxies for PPP based on longer prescription histories result in stronger IVs, partly offsetting the effect on power of smaller sample sizes.

Conclusion

Irrespective of sample size, the PPP IV approach leads to less biased estimates of treatment effectiveness than conventional multivariable regression adjusting for known confounding only. Particularly for smaller sample sizes, we recommend constructing PPP from long prescribing histories to improve statistical power.

Keywords: simulation study, comparative effectiveness research, instrumental variables, Physician's prescribing preference

7.3 Introduction

As a source of natural variation, PPP has been increasingly used as an IV in CERs (Brookhart and Schneeweiss, 2007). Multiple simulation and applied studies have discussed the use of PPP in comparing the effectiveness of two drug classes. In many recent applied papers about PPP IV, they have large sample sizes of around 30,000 (Kuo et al., 2012, Davies et al., 2020, Kollhorst et al., 2016, Taylor et al., 2017). However, in many contexts the sample size will be smaller, for example, Nelson and colleagues conducted a PPP study of HIV using a sample size of less than 2000 (Nelson et al., 2013). Smaller sample sizes are likely to occur in studies of rare outcomes or where drugs have only recently become available (e.g., in a single administrative area).

Boef and colleagues argued that the sample size put limits on the performance of IVs (Boef et al., 2014). Further, they concluded that the bias in IV estimates relative to conventional approaches (e.g., OLS) is determined both by the strength of the IV as well as the strength of unmeasured confounders. With an aim to widen the applicability of PPP IV, we test the performance of the method in moderate and small sample sizes using a simulation study.

7.4 Method

7.4.1 Statistical analysis approaches

In order to be comparable with OLS, we use 2SLS as the main statistical method to generate the IV estimates of treatment effectiveness. Despite the fact that 2SLS may cause model misspecification for binary outcomes and treatment, the 2SLS is the most common method and a common starting point for the IV method (Zhang et al., 2018). In addition, in many settings, when the outcome is not rare, 2SLS generates similar estimates to non-linear two stage regression (prevalence between 1.5% to 50%) (Ionescu-Ittu et al., 2009).

A summary of how performance was assessed is shown below (See Table 18). We use percent bias to assess the performance of PPP IVs for different levels of unmeasured confounding. The strength of IV is calculated as the F-statistics of the first stage. We use the coverage rate to compare the stability of OLS and 2SLS at the different levels of unmeasured confounding.

Measurement	Calculation
Percent bias	$\frac{true\ Risk\ Difference - estimated\ Risk\ Difference}{true\ Risk\ Difference}$*100%
Coverage rate	% of iterations when 95% CI includes the true risk difference across 1000 simulations
F-statistics of the first stage regression	F-statistics $= \frac{Sum\ of\ squares\ for\ Model / Degrees\ of\ Freedom\ For\ Model}{Sum\ of\ Squares\ for\ Error /\ Degrees\ of\ Freedom\ for\ Error} = \frac{Mean\ of\ Squares\ for\ Model}{Mean\ of\ Squares\ for\ Errors}$

Table 18. Measurement of performance.

7.4.2 Simulation design

7.4.2.1 Study population

For the moderate sample size study, we set the number of physicians to 80, the lower bound of the number of patients/physicians was 10 and the upper bound of patients/physician was 50. The overall sample size in this case is 2453. For small sample size study, the number of physicians is set to 20. The sample size is 620. The prevalence of outcomes varies between 20% to 60%.

7.4.2.2 Treatment and outcome

In this paper, we focus on scenarios where the treatment and outcome are both binary.

The formula for the probability of being prescribed a certain treatment (X = 1) and the probability of the outcome of interest (Y = 1) are listed below:

$$Prob(X = 1) = \alpha_0 + \alpha_z PPP + \alpha_1 X_1 + \gamma_x X_2 + \alpha_3 X_3$$

$$Prob(Y = 1) = \beta_0 + \beta_x * Prob(X = 1) + \beta_1 X_1 + \gamma_y X_2 + \beta_3 X_3$$

PPP stands for IV. We set PPP 70% of chance equals to 1, 30% of chance equals to 0. This imbalance reflects a common situation that treatment providers tend to prefer one type of treatment than another (perhaps based on following clinical guidelines). X_1 is a binary covariate, and X_2, X_3 are continuous covariates. We assume X_1 and X_3 are measured covariates and X_2 is an unmeasured covariate. In the data generation process, X_1 follows binominal distribution, X_2 and X_3 follows the normal distribution. These are implemented using R functions rbinom and rnorm (please see the R code in supplementary material for full details). α_z controls the strength of association between the instrumental variable and exposure. The PPP is the 'true' prescribing preference that in practice is a latent variable and is a binary variable. The parameter values for the data generation process are listed in equation (1) and (2).

The focus of this study is to investigate the impact of unmeasured confounding. Therefore, we keep α_z=0.4 to ensure the IV strength is fixed. The parameter value for treatment in equation (2) is 0.1 and this represents the 'true' risk difference between the two treatments. β_x is the observed estimate of this risk difference.

$$[Prob(X = 1)] = \alpha_z PPP + 0.053X_1 + 0.1\,X_2 + 0.02X_3 \quad (1)$$

$$[Prob(Y = 1)] = 0.10 * treatment + 0.04X_1 + \gamma_2 X_2 + 0.01X_3 - 0.01 \quad (2)$$

Drawn from the existing literature (Davies et al., 2013a, Davies et al., 2020, Brookhart and Schneeweiss, 2007, Taylor et al., 2017), we constructed the proxies for PPP mainly based on the prescription history. The prior 1 to prior 4 prescriptions are investigated in this study. The prior 1 prescription is the most recent prescription made by the same physician. Likewise, the prior 2 prescription is prior 2 prescriptions from the same physician and the same for prior 3 and prior 4 prescriptions. For example, possible values for prior 4 prescriptions are 0,1,2,3,4. The proportional PPP is the number of certain treatment (X=1) divided by the number of all prescriptions made by this physician (See Equation 9).

$$Proportional\ PPP = \frac{Number\ of\ drug\ A\ made\ by\ one\ physician}{Number\ of\ all\ prescriptions\ made\ by\ the\ same\ physician}$$

Equation 9. Calculation of the proportional PPP

All analysis is done in R studio using R version 3.6.1. The R code that generates the simulated datasets and the regression models can be seen in GitHub link provided in supplementary material.

7.5 Results

Figure 10 presents the percent bias of the 2SLS and OLS in moderate and small sample sizes. OLS is subject to unmeasured confounding bias. In the case of a lower unmeasured confounding level, the 2SLS is more biased than OLS. The advantage of 2SLS appears after the level of the unmeasured confounding increases. The sample size does not influence the percent bias in general.

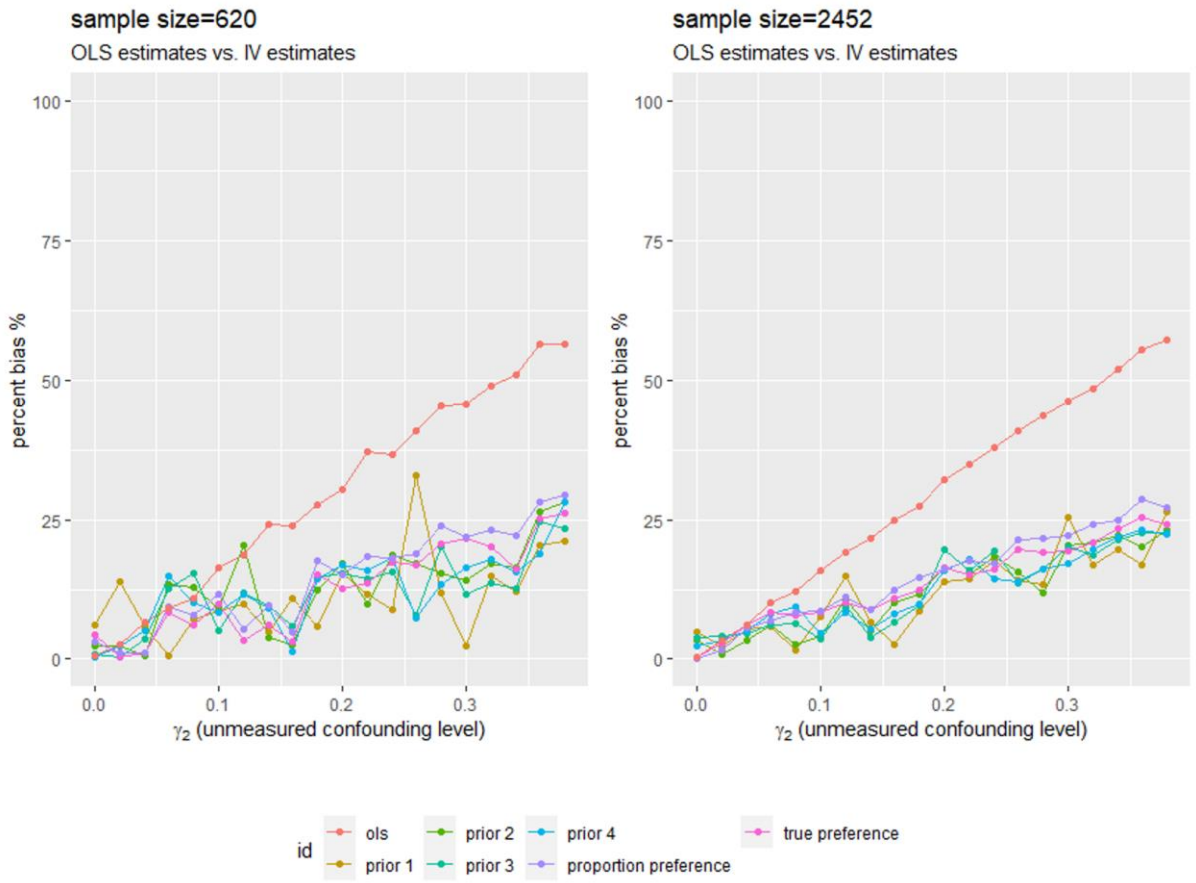


Figure 10. Percent bias of 2SLS and OLS.

The coverage rate shows that that the 2SLS covers nominal 95% while the coverage rate of OLS drops dramatically in both sample sizes (see Figure 11). This can be explained by Equation 10 where the difference between the variances of OLS estimates and variances of IV estimates is defined by the value of the correlation between the treatment and the IV ($\rho_{X,Z}$) (Martens et al., 2006). The value of correlation between the treatment and IV are no larger than 1 which make the variance of IV larger than that of the OLS.

$$var\left(\hat{\beta}_n^{IV}\right) = \frac{\sigma_{Y,X}^2}{n\,\sigma_X^2 \rho_{X,Z}^2}$$

$$var\left(\hat{\beta}_n^{OLS}\right) = \frac{\sigma_{Y,X}^2}{n\,\sigma_X^2}$$

Equation 10. Variance of IV estimate (2SLS) and OLS estimate. $\sigma_{Y,X}^2$*:The residual variance of the outcome after adjusting the treatment (X);* σ_X^2*: The variance of the treatment;* $\rho_{X,Z}^2$*: The correlation between the treatment(X) and the instrumental variable (Z).*

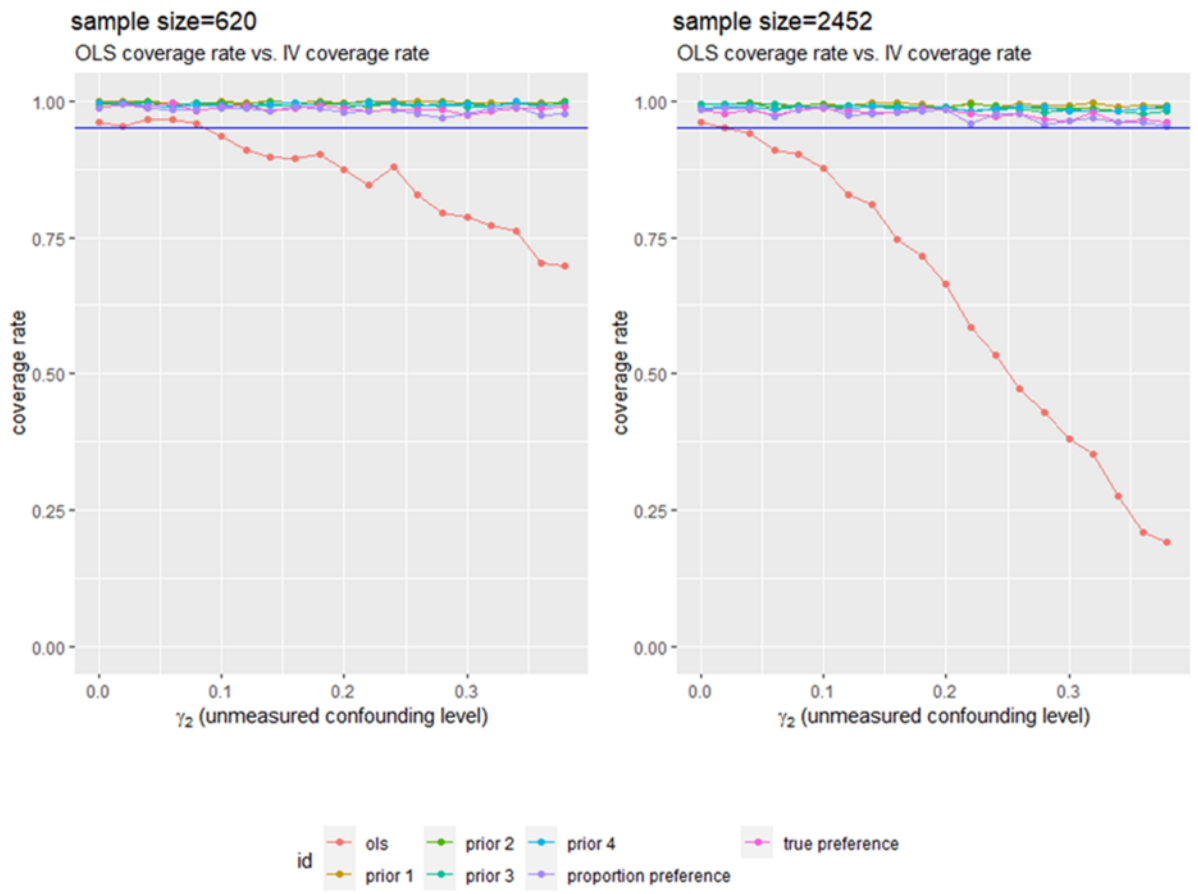


Figure 11. Coverage rate across 1000 times simulation. The blue intercept line represents the nominal 95%

The strength of the IV increases, and the p value of the 2SLS estimate decreases, as the number of previous prescriptions used in the PPP construction increases (See Figure 12). The level of unmeasured confounding does not influence these results. However, the strength of IV decreases noticeably when the sample size decreases. The relation between the F-statistics, sample size and the correlation between the treatment and the IV is shown in Equation 11. From the simulated data, ρ_{zx} does not change much in these two cases (around 0.14 to 0.15) indicating that the strength of the association between the exposure and IV does change. Rather, it is the sample size that decreases the F-statistics and makes the IV weaker (Martens et al., 2006). The p values of 2SLS in N=620 sample are consistently larger than that of N=2,452 which means the statistical power of 2SLS is limited by the sample size.

$$F\ statistics = \frac{\rho_{ZX}^2(n-2)}{1-\rho_{ZX}^2}$$

Equation 11. σ_X^2: The variance of the treatment; $\rho_{Z,X}^2$: The correlation between the treatment(X) and the instrumental variable (Z).

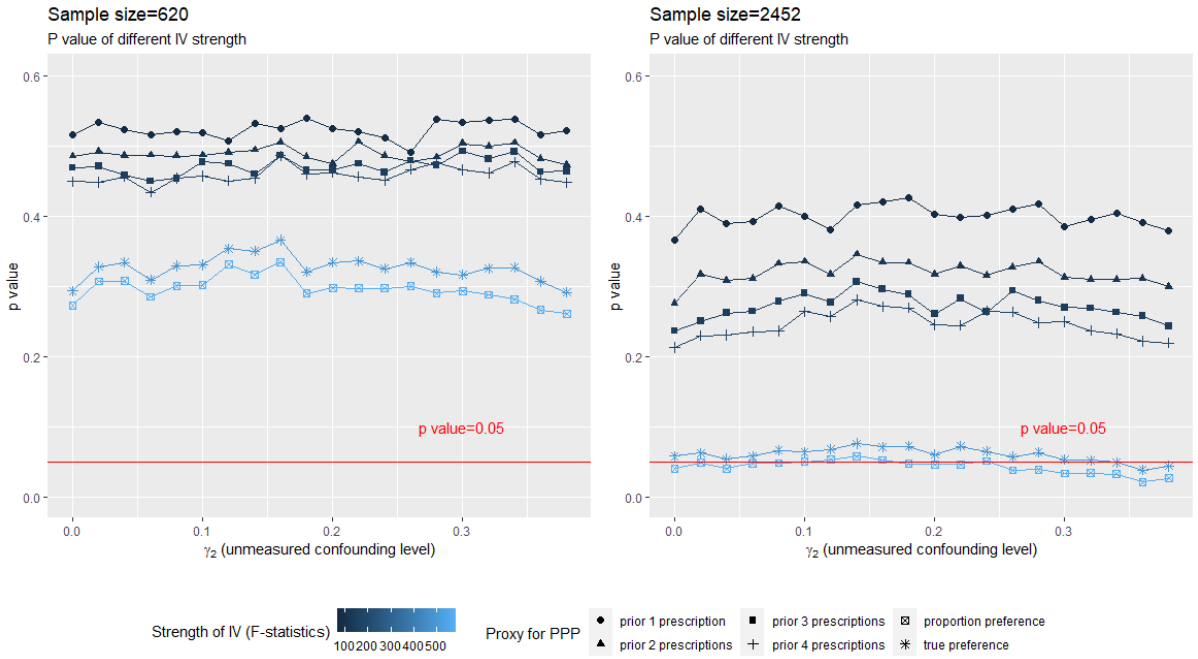


Figure 12. p values of OLS and 2SLS estimates.

As an IV, the true preference (PPP in the model (1)) also shows a strong ability to reduce the unmeasured confounding bias. The F-statistics of true preference reaches 500 which is much higher than all proxies mentioned above which align with the finding from Ionescu-Ittu et al. (Ionescu-Ittu et al., 2009) that the true preference has the smallest variance. The p values for 2SLS estimates are close to conventional statistical significance (p value <0.05). The bias-variance trade-off for IV methods also exist for the 'true preference' but not as critical as for the proxy PPP indicating stronger instrument reduces the variance of instrumental variable estimates (Ionescu-Ittu et al., 2012). For the reason that time cannot be simulated, we test the time-fixed proxy for PPP (proportional preference). It turns out that the proportional preference is the strongest IV among these proxies. It is associated with the smallest p value which leads to the 2SLS estimates having small p values.

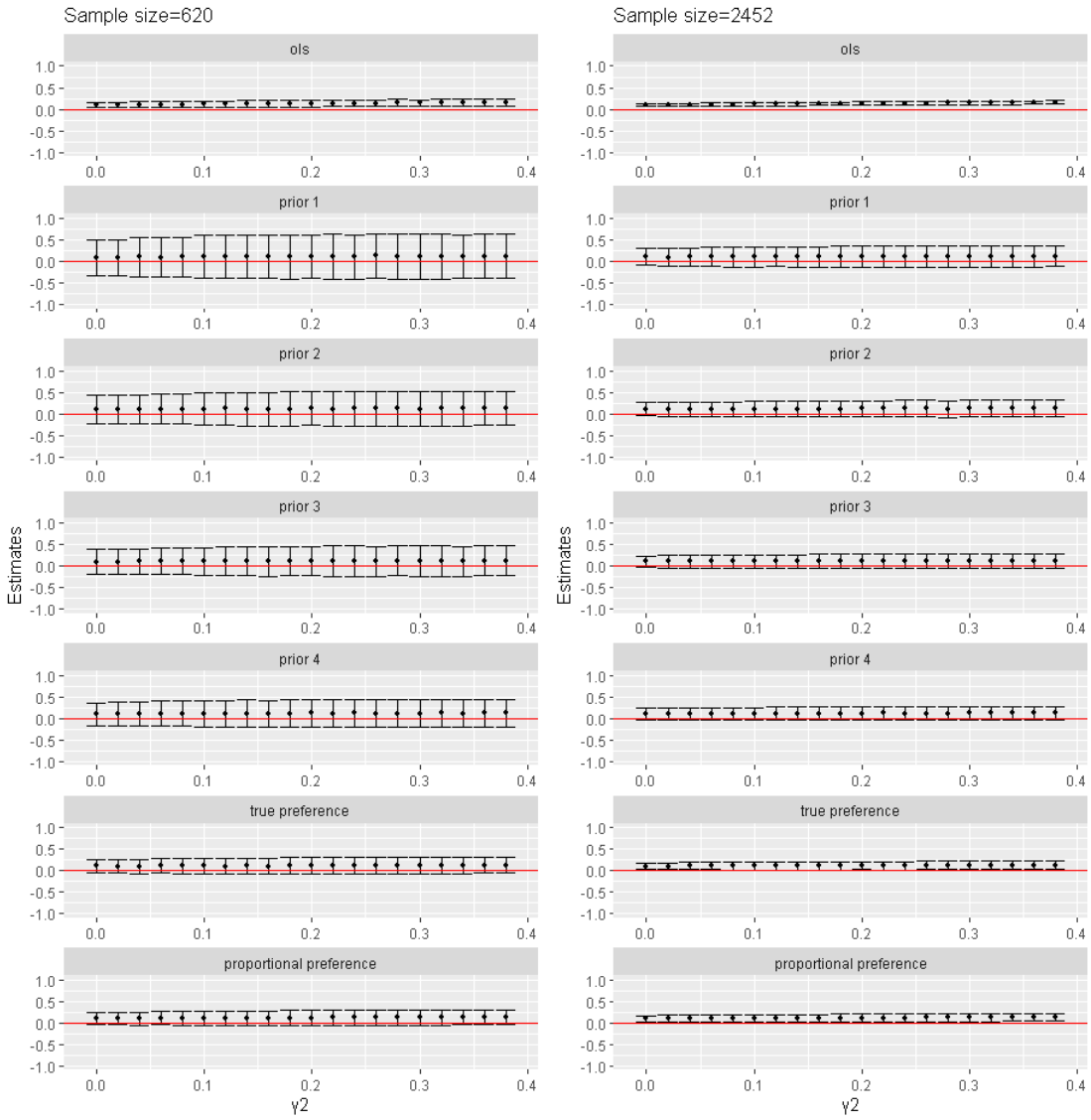


Figure 13. 95% Confidence intervals of OLS and 2SLS estimates (95% CIs are calculated using cluster robust standard errors). The red line represents the null hypothesis.

As summarised in Figure 13, 95% CIs of IV estimates narrow as the strength of instruments increases (from prior 1 prescription to the proportional preference). As discussed above, the 95% CIs of OLS estimates are narrower than for 2SLS, and this is also shown in Figure 4. It can also be seen that the OLS estimates are severely biased when the unmeasured confounder covariate parameter (γ_2) is set at a high level. Although the IV estimates are generally less precise, it is feasible when the IV is strong enough that an IV estimate can achieve statistical significance while at the same time reducing the influence of unmeasured confounding bias.

7.6 Discussion

The sample size limits the performance of IVs (Boef et al., 2014, Martens et al., 2006). A straightforward explanation for this is that smaller sample sizes make it harder for the IV to meet the relevance assumption. In real life CERs, sample sizes are often large enough which can avoid such pitfalls, but when the outcome of interest is rare, or a drug has only recently become available the corresponding CERs will have smaller sample sizes. This simulation aimed to test the performance PPP IV on the different level of unmeasured confounding level and generate supporting evidence that the PPP IV can perform well in reducing bias in studies of moderate or small sample sizes.

Our results show that 2SLS does reduce the unmeasured confounding to a considerable extent compared to conventional analyses even in a small sample size. At the same time, the standard deviation of 2SLS estimates is generally many times larger than OLS and the confidence interval wide and crossing the null hypothesis. However, if the instrumental variable is strong enough, 2SLS estimates could be statistically significant. In terms of reducing bias, the sample size is not the determinant as it does not impact percent bias. Nevertheless, the sample size does limit the statistical power. In a smaller sample size, the difference of percent bias from 2SLS and OLS can still be an indicator to see if an unmeasured confounding is a major problem although the weak statistical power makes the 2SLS estimate less useful.

The results of this simulation study show that using PPP as an IV is effective at minimising bias caused by unmeasured confounding relative to only adjusting for measured confounding in CER. The PPP is a latent variable that cannot be measured directly using routinely collected data (Brookhart and Schneeweiss, 2007). Our results show that increasing the number of previous prescriptions used in constructing the PPP leads to power gains which could be particularly important for studies with small or moderate sample sizes. It is worth noting that using PPP with only one previous prescription is a popular strategy in the applied literature. According to our results, prior 2, prior 3, prior 4 and the proportional IV performs better than prior one since the IV strength increases as we account for longer prescription history. It is worth pointing out that by using a larger history in calculating PPP, this implicitly assumes that a physician's preference does not change over time. This can be empirically tested using study data.

Baiocchi and colleagues suggest that researchers should consider the necessity for using IV method to account for unmeasured confounding. If the unmeasured confounding is small, IV methods may not be necessary (Baiocchi et al., 2014). I support this conclusion with my simulation results. According to the figures, it is quite noticeable that there a threshold where the per cent bias of conventional methods become larger than that of IV methods. If we conduct IV methods at those points, the IV estimates may not be reliable, especially when we use 2SLS (see the 2SLS vs. OLS figure when the γ_2 equals 0) to compare with the OLS.

The limitation of this study mainly rests on the simplicity of the design. By moderate sample size, we used approximately 2500 which is derived from a research study the authors led on investigating prescribing for alcohol dependence in Scotland. Also, I reviewed the sample sizes that in the current CERs papers that focus on the PPP IV and found that most of them are above 10,000. I did not consider survival analysis including censored outcomes (Tchetgen et al., 2015) or non-linear two-stage approaches, like two-stage predictor substitution and two-stage residual inclusion, in the simulation design. Finally, I need to emphasise that an essential limitation of studying the time-based proxies for IVs is that the time cannot be truly simulated as all data generated at the same time. The prior 1,2,3,4 prescription proxy estimates are based on real time in applied studies. Strictly speaking, this simulation demonstrates valid proxies for PPP IV, rather than the "true" prior 1,2,3,4 prescriptions as proxies.

7.7 Conclusion

Using PPP as an IV for CER is less biased than conventional approaches and can achieve adequate statistical power in smaller sample sizes if the IV strength is high enough. If it can be assumed that a physician's prescribing preference does not change over time, we recommend constructing PPP using entire prescribing history to gain power.

Chapter 8. Comparing the performance of two-stage prediction substitution and two-stage residual inclusion methods when using physician's prescribing preference as an instrumental variable in comparative effectiveness research.

8.1 Publication details

This chapter had been submitted for publication in Journal of comparative effectiveness (under review).

8.2 Abstract

Background

Instrumental variable (IV) methods are widely used to address unmeasured confounding concerns in comparative effectiveness research (CER). When the outcome variable is binary, two-stage residual inclusion (2SRI) and two-stage prediction substitution (2SPS) are the most commonly used two-stage non-linear IV methods. To date, CER studies focus on the comparison between conventional methods (e.g., multivariable regression, propensity scores) and IV methods in terms of their capability of reducing unmeasured confounding bias. However, the concern of noncollapsibility effects in non-linear settings has been overlooked.

Objective

The first objective is to compare the performance of 2SRI, 2SPS with the multivariable generalised linear model (GLM) in terms of the reducing unmeasured confounding bias. The second objective is to demonstrate the ability of 2SRI and 2SPS in alleviating unmeasured confounding when noncollapsibility exists.

Methods

This study comprises a simulation study and an empirical example from a real-world UK population health data set (Clinical Practice Research Datalink). The IV used is based on physicians' prescribing preferences (defined by prescribing history).

Results

We found the percent bias of 2SRI in terms of treatment effect estimates to be lower than GLM and 2SPS and was less than 15% in most scenarios. Further, 2SRI was found to be robust to mild non collapsibility with the percent bias less than 50%. As the level of unmeasured confounding increased, the ability to alleviate the noncollapsibility decreased. Strong IVs tended to be more robust to noncollapsibility than weak IVs.

Keywords: noncollapsibility, unmeasured confounding, instrumental variables, Physician's prescribing preferences, two-stage prediction substitution, two-stage residual inclusion.

8.3 Introduction

In order to address unmeasured confounding bias concerns in observational CERs, the IV approach is widely used. In this approach, the most commonly used estimation method is the 2SLS which consists of two stage linear regression. The 2SLS estimator is normally consistent when the outcome measure is represented as a numerical variable (Palmer et al., 2017). However, if one requires to estimate the treatment effect using an OR for a binary outcome, the method needs to be adapted to the non-linear setting. One such approach is two-stage predictor substitution (2SPS). The first stage regression of 2SPS is treatment regressed upon the covariates; the second stage is the outcome regressed upon predicted results from the first stage together with covariates.

Another non-linear method, two-stage residual inclusion (2SRI), has the same first stage regression as 2SPS, but use the residuals from the first stage as an additional covariate in the second stage. It was firstly introduced by Hausman (Hausman, 1978) in order to test endogeneity in the linear context. Currently, there are simulation studies (Terza et al., 2008a, Terza, 2018, Cai et al., 2011) as well as real-world studies that provide evidence for the 2SRI being generally less biased than 2SPS when estimating a treatment effect in the presence of unmeasured confounding. However, unlike risk difference, odds ratio, is not collapsible which means that it cannot always be expressed as the weighted average of stratum-specific OR. This characteristic also refers to noncollapsibility (Schuster et al., 2021, Pang et al., 2016, Greenland et al., 1999). For example, if one adjusts for covariates that are not associated with both outcome and treatment in a logistic regression model (i.e., not a true confounder), the adjusted OR may differ from the unadjusted OR. Therefore, in such contexts the difference between adjusted and unadjusted logistic regression consists of is made of two parts: confounding effect and noncollapsibility effect.

In terms of using PPP as an IV and applying the 2SRI method, Koladjo et al. concluded that 2SRI is less biased than IV based GMM (Koladjo et al., 2018) in estimation of treatment effect. It is widely acknowledged that 2SPS is not superior to 2SRI in terms of dealing with endogeneity in health research (Cai et al., 2011, Terza et al., 2008a). However, there are also studies indicating that 2SRI produces biased estimates of average treatment effect (ATE) and local average treatment effect (LATE), compared with 2SLS (Basu et al., 2018). In this study, I focused on the non-linear settings. For the conventional approaches which do not account for the unmeasured confounding issue, I chose the generalised linear model (GLM) as it is a one of the most intuitive approaches in non-linear settings. There are two objectives in this study: 1) Compare 2SRI and 2SPS with the generalised linear models (GLMs), which can only adjust for measured confounders, in a drug comparison simulation study using physician's prescribing preference as instrumental variable in the presence of unmeasured confounding bias; 2) Test the robustness of 2SRI to noncollapsibility, using simulated data and real-life data from a real-world UK population health data set (CPRD).

8.4 Method

8.4.1 Data generating process.

In order to construct an observational CER, I set the total number of physicians as 80. The patients per physicians is in range from 10 to 50. The simulated data consists of 2,442 records (n=2,442). X_1 and X_2 are the measured confounders. 'un' is the unmeasured confounder. The R code for constructing the treatment (X) and the outcome (Y) is listed in Box 1 (for the details see GitHub link provided in supplementary material).

Research objective 1
X_1 <- rbinom(n, 1, 0.6) X_2 <- rnorm(n,13,2) un<- rnorm(n, 1,2) x<- 4*PPP+0.2* X_1 + 0.53*X_2-3.5+1.1*un ptxA= exp(x)/(1+exp(x)) txA <- rbinom(n, 1,ptxA) y<- -0.9*txA-0.02*X_1 -0.6*X_2 - γ_2*un pout <- exp(y)/(1+exp(y)) out <- rbinom(n, 1,pout)
Research objective 2
X_3 <- rnorm(n,10,1) x<- 4*PPP+0.2*X_1 + 0.53*X_2 -3.5+1.1*un+0*X_3 ptxA= exp(x)/(1+exp(x)) txA <- rbinom(n), 1,ptxA) y<- -0.9*txA-0.02*X_1 -0.6*X_2 -γ_2*un +γ_3*X_3

Box 1. R code for the data generation

X_3 in research objective 2 is the variable that induces the noncollapsibility effect which is based on a scenario that the variable is associated with the outcome but not associated with the treatment (Schuster et al., 2021). X_3 is formed with a mean value of 10 and 1 as standard deviation to ensure an adequate non-collapsibility effect. I used the γ_2 (ranges from 0 to 1.9) to control the unmeasured confounding level and γ_3 (ranges from 0 to 0.95) to control the level of noncollapsibility effect. The PPP IV is formed by the prior n prescription of drug A and divided by n prescribed by the same physician. The strength of IV is tested using the F-statistics. All assumptions of a valid IV are assumed to be met in this simulated dataset.

8.4.2 Study design

According to recent studies, 2SRI is sensitive to the choice of residuals (Basu et al., 2018). In this study, I selected Pearson residuals to be used in the second stage of regression after initial analysis using raw residuals (results from raw residuals are extremely biased and not shown in this thesis). Percent bias and coverage rate are used to measure the performance of the estimation methods (See Table 19). In order to obtain more precise estimates, each simulation is run for 1000 times. All simulations and statistical analyses are conducted using R version 4.1.1.

Measurement	Calculation
Percent bias (in GLM, 2SRI, 2SPS)	$\frac{true\ odds\ ratio - estimated\ adds\ ratio(from\ GLM,\ \ 2SRI\ and\ 2SPS)}{true\ odds\ ratio}$*100%
Coverage rate	% of iterations when 95% CI includes the true OR across 1000 simulations
F-statistics of the first stage regression	F-statistics $=\frac{Sum\ of\ \ squares\ \ for\ Model/Degrees\ of\ Freedom\ For\ Model}{Sum\ \ of\ Squares\ \ for\ Error/\ \ \ Degrees\ of\ Freedom\ for\ Error}$ $=\frac{Mean\ \ of\ \ Squares\ \ for\ Model}{Mean\ of\ Squares\ for\ Errors}$

Table 19. Measurement of performance

8.4.3 Design of the empirical illustration

An empirical example is presented in this section to demonstrate the ability performance of 2SRI in dealing with the accounting for a noncollapsibility effect. The data used in this section is the study cohort in the comparative effectiveness research section in Chapter 5 (see section 5.5.3.3). It is a CER which compares the effectiveness of acamprosate and disulfiram in reducing the risk of alcohol use disorder (AUD) hospitalisations in England using data from CPRD. Unlike adjusted logistic regression, the inverse propensity score weighting (IPSW) method using stabilised weights is considered to estimate the marginal treatment effect (MTE) and is free from the impact from noncollapsibility. Therefore, the noncollapsibility is usually quantified as the difference between multivariable logistic regression and IPSW adjusted results, while the confounding bias is quantified using the difference between the univariable logistic regression and the IPSW adjusted results (Pang et al., 2016, Schuster et al., 2021). The IV used in this case is the proportion of acamprosate among the last year prescriptions.

8.5 Results

8.5.1 Research objective 1: Assessing the ability of 2SRI and 2SPS of alleviating the unmeasured confounding bias.

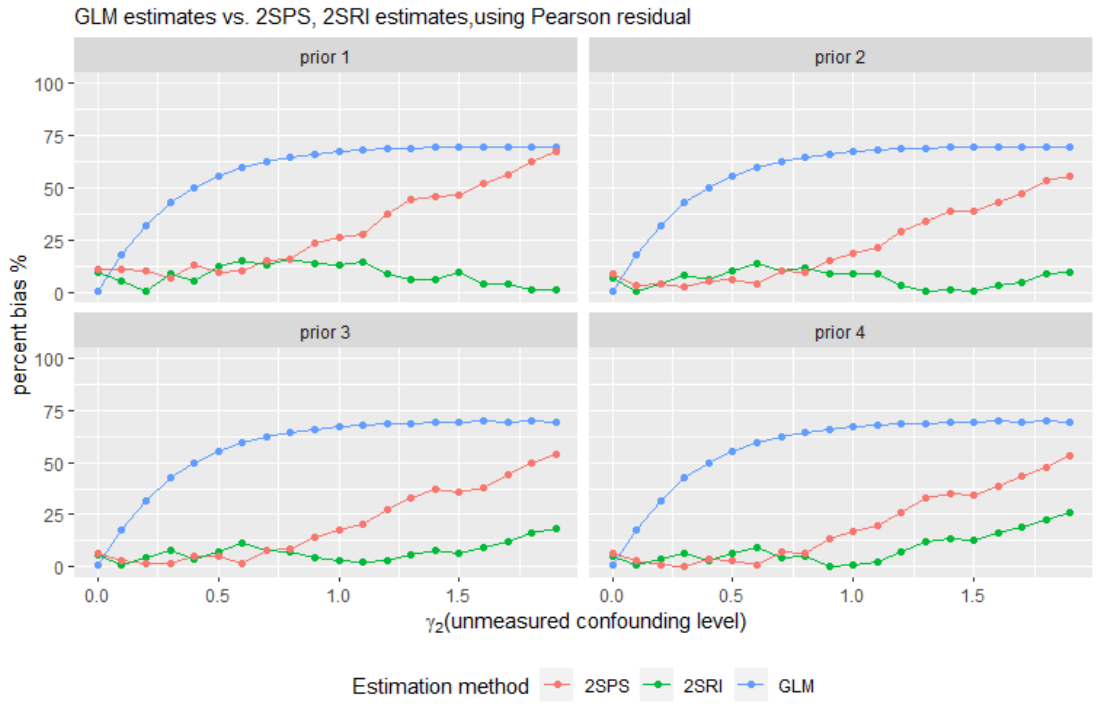


Figure 14. Estimates from GLM, 2SRI and 2SPS.

Figure 14 shows that 2SPS is consistently more biased than 2SRI (γ_2 more than 0.75). When γ_2 is less than 0.75, the 2SRI estimates are not always less biased than 2SPS, but the percent bias is consistently at a low level (below 12.5%). The percent bias from 2SRI does not always inflate as the unmeasured confounding level rises; in the case of prior 1 and prior 2 as IV, I observed a rather low percent bias (less than 12.5%) for 2SRI throughout the range of γ_2 values from 0 to 1.9. The estimates from GLM are only at a low level when the unmeasured confounding level is small.

Despite the point estimate deviating from the 'true' OR, the coverage rates of 2SPS are around 95% most of cases. When the IV strength increases (F-statistics from 105 to 250), the coverage rates from 2SRI are more likely to achieve 95% (see Figure 15).

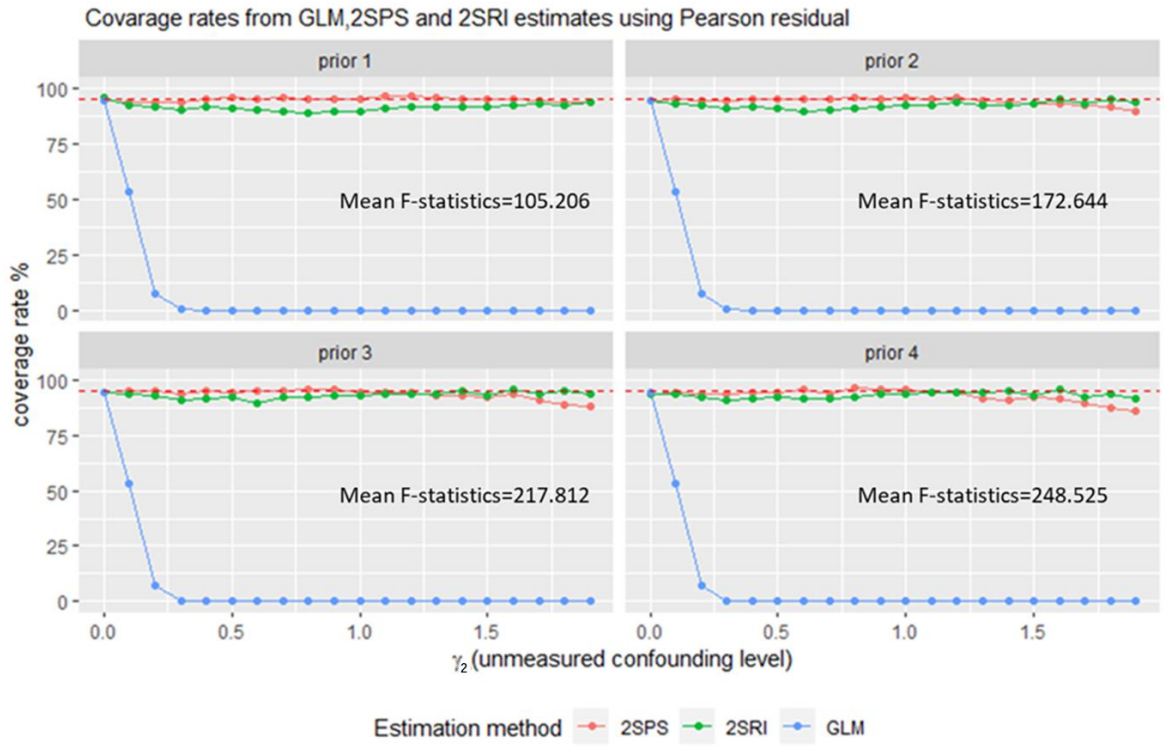


Figure 15. Coverage rates from 2SRI, 2SPS, GLM. Red dash line represents the 95% nominal.

8.5.2 Research objective 2: Assessing the ability of 2SPS and 2SRI of alleviating noncollapsibility.

This simulation tested whether the 2SRI or 2SPS estimates are able to reduce unmeasured confounding bias. However, the percent bias from the research objective 1 is free from the noncollapsibility effect. Research objective 2 is to assess the performance of 2SRI and 2SPS with the existence of unmeasured confounding as well as the noncollapsibility effect. I minimised the unmeasured confounding effect by selecting two scenarios where γ_2 equals 1.0 and 1.5 where 2SRI is generally unbiased (percent bias less than 10%) against the unmeasured confounding according to the results from the research objective 1. The results are shown in Figure 16.

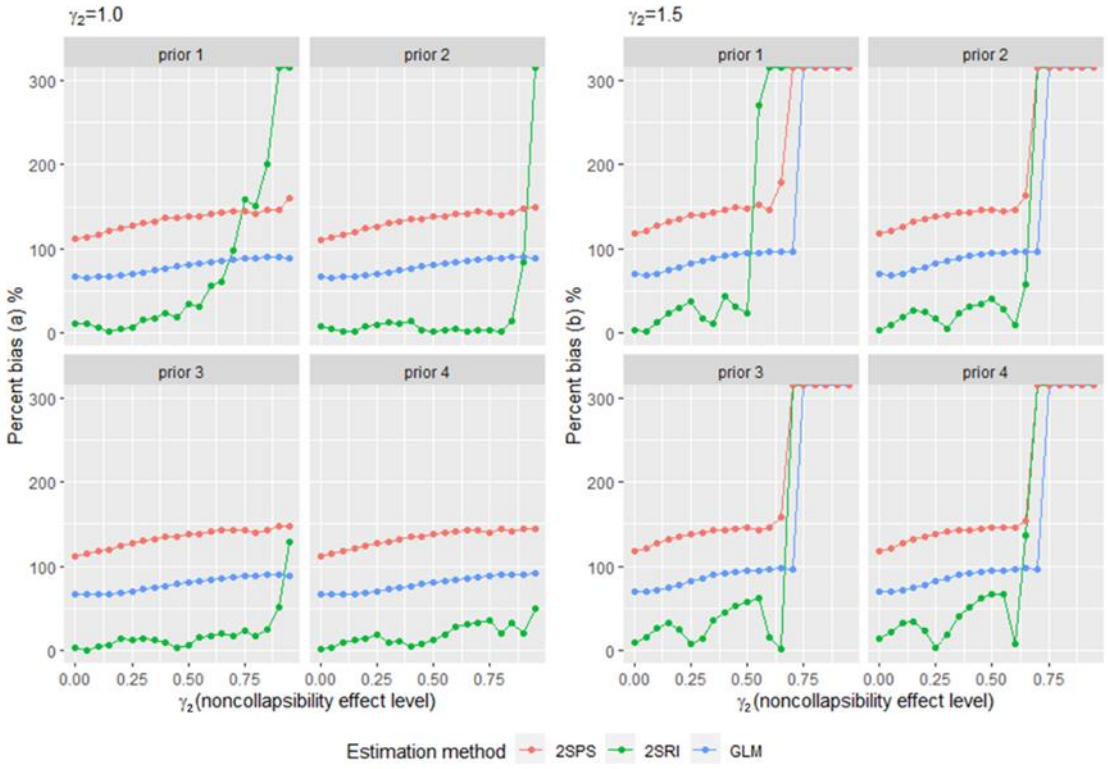


Figure 16. Percent bias of 2SRI, 2SPS, GLM when there is noncollapsibility. Percent bias (a) represents percent bias of the estimate when γ_2 equals 1.0. Percent bias (b) represents percent bias of the estimate when γ_2 equals 1.5.

It can be seen from Figure 16 that the 2SRI is generally less biased than 2SPS and GLM when the noncollapsibility effect is not severe. When the γ_2 equals 1.0, the percent bias of 2SRI is at a low level at the beginning but rise dramatically when the noncollapsibility effect increases. Same trend is found in GLM and 2SPS. When the γ_2 equals 1.5 (the unmeasured confounding effect at higher level), the percent bias of 2SRI fluctuated below 50% and hits a high level as γ_3 increases. The threshold of γ_3 where the percent bias (b) of 2SRI becomes extremely biased is smaller than percent bias (a) of 2SRI. The percent bias of 2SPS and GLM exceeds 100% when γ_3 is at low level in both scenarios indicating they are less robust to noncollapsibility. Note that for 2SPS and GLM, the percent bias is high even when γ_3 equals 0. The coverage rates are presented in Figure 17. The coverage rates of 2SRI are around 95% when the γ_3 equals 1.0 and drop when the noncollapsibility effect increases. Note that, the coverage rate (b) increases after a certain point because the confident interval of 2SRI estimate is extremely wide, where the estimates of 2SRI become biased.

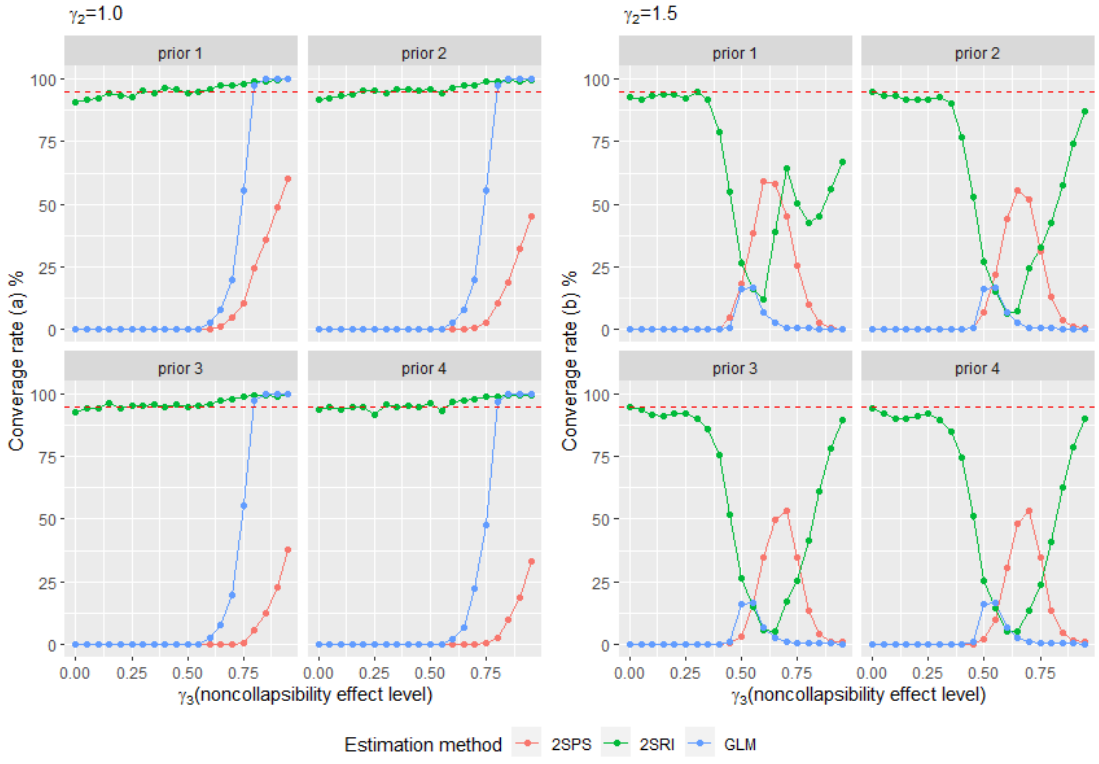


Figure 17. Coverage rate of GLM, 2SRI and 2SPS. Coverage rate (a) represents coverage rate of the estimate when γ_2 equals 1.0. Coverage rate (b) represents coverage rate of the estimate when γ_2 equals 1.5.

Empirical illustration using CPRD Data

Model	β'	β	β^{IPSW}	β^{2SRI}	$\beta^* = \beta - (\beta - \beta^{IPSW})$	Percent difference $= \frac{\lvert\beta^{2SRI} - \beta^*\rvert}{\beta^*} * 100\%$
Model 1	0.168	0.211	0.184	0.152	0.238	33.9%
Model 2	0.168	0.439	0.420	0.326	0.458	28.8%
Model 3	0.439	0.452	0.379	0.608	0.525	13.6%

Table 20. Results from empirical case study. β' represents the estimate from unadjusted logistic regression. β represents the estimate from adjusted logistic regression. β^{IPSW} represents the estimate from the IPSW using stabilised weight. β^{2SRI} represents the estimate from adjusted 2SRI.

As it mentioned in the section 8.4.3, the empirical illustration of research objective is based on a CER that compares the effectiveness of acamprosate and disulfiram in preventing AUD hospitalisation. The confounders adjusted in the models are Charlson Comorbidity Index (CCI), and 'prescription year'. CCI is associated with the outcome, but not associated with the treatment. 'Prescription year' is associated with both treatment and outcome. The presence of confounding bias and/or noncollapsibility is defined by adjusting or not adjusting these confounders. Model 1 is a logistic regression with or without adjusting CCI. Model 2 is a logistic regression with or without adjusting 'prescription year'. Model 3 is a logistic regression adjusts 'prescription year', with or without adjusting CCI. The results of the case study are shown in Table 20. Since the CCI is associated with outcome but not associated with the treatment, the difference between the coefficient from β' and β should be totally due to the noncollapsibility. According to the difference between the β^{IPSW} and β (0.184-0.211= -0.027) and the difference between β' and β^{IPSW} (0.168-0.184=-0.016), the confounding effect is less significant. Excluding the noncollapsibility from β is calculated as: 0.211- (0.184-0.211) =0.238. But In Model 2, the estimates that excludes the noncollapsbility should be around β : 0.439- (0.420-0.439) =0.458. In Model 3, the noncollapsibility is calculated as 0.379-0.452=-0.073. The estimate of treatment from the adjusted model that excludes the noncollapsibility is calculated as 0.452- (-0.073) =0.525. β^{2SRI} is close to 0.525 and with a smallest percent difference compared with Model 1 and Model 2. In Model 1 and Model 2, 2SRI does not show much ability to remove the noncollapsibility from the adjusted model. However, results from the Model 3 echoes the simulation in the research objective 2 that the 2SRI can alleviate noncollapsibility when true confounders are adjusted for in the model. Note the percent difference in Table 20 may be partly due to the residual unmeasured confounding in the observational studies.

8.7 Strengths and Limitations

To my knowledge, this study is the first simulation study to discuss the 2SRI's robustness to the noncollapsibility effect. In addition to the simulated data, I demonstrated an empirical example from CPRD to illustrate the noncollapsibility effect is common in logistic regression and can cause misleading results in causal inferential studies. One limitation of this study is the simplicity of the design. I did not consider more covariates, or more than one IV, or another forms of residuals that used in 2SRI. Another limitation I assumed the IV assumptions are met in all simulated scenarios.

8.6 Discussion

Results from the simulation study show that the percent bias of 2SRI is less than 15% in most scenarios while the percent bias of 2SPS reaches 50%. This echoes the findings from Cai et al.(Cai et al., 2011), who found that two-stage logistic regression 2SRI is asymptotically unbiased when the unmeasured confounding effect is not severe. However, my results are inconsistent with their conclusion that percent bias of 2SRI tends to rise as the unmeasured confounding level increases. My findings indicate that the percent bias of 2SRI fluctuates when the unmeasured confounding is moderate but does not increase monotonically following the increase of an unmeasured confounding effect. According to the results, the IV strength does not affect the consistency of 2SRI estimates.

I discussed the potential bias from the noncollapsibility on the 2SRI estimate using a simulated scenario where a covariate is associated with outcome but not associated with the treatment. The results indicate that percent bias of 2SRI has potential to be robust with minor or moderate noncollapsibility effect, and robustness is associated with the level of unmeasured confounding effect. I can see from Figure 16 that robustness is shown to be more resilient when the unmeasured confounding is smaller in magnitude (γ_2 equals 1.0). It is also reflected in the case study that 2SRI provides a close estimate to the unadjusted logistic regression when a variable that should not be adjusted appears in a model and brings noncollapsibility. However, for 2SPS and 2SRI, the noncollapsibility leads to major distortions on the estimates.

In this study, 2SRI is shown to be superior to 2SPS. There are studies that argue the consistency of 2SRI estimate is associated with the collapsibility of the model. Normally, the 2SRI estimator is consistent when the model is collapsible, for example in the addictive hazards models (Wang et al., 2017a). However, my results show that the 2SRI estimate is not certainly biased with noncollapsibility effect. Despite my results supporting the preference of 2SRI, the non-linear extension of 2SRI and 2SPS in the binary exposure are not studied adequately. Wan et al. proved the consistency of 2SRI estimates as the same time pointed out that the original framework proposed by Terza is used for the continuous treatment variable (Wan et al., 2018). Further theoretical and methodological research is needed for 2SRI used for binary treatment.

8.8 Conclusion

The findings of this simulation study show that 2SRI performs unbiasedly in non-linear models when conducting comparative effectiveness research. Further, the results show that 2SRI is more likely to alleviate noncollapsibility when unmeasured confounding effects are at lower levels.

Chapter 9. General discussion

In this chapter, I start by restating the research objectives of the thesis before providing a summary of the previous chapters and how they addressed the research objectives. Then, I summarise the key contributions I have made in context to the existing literature, before going on to discuss the strengths and limitations of my thesis. I conclude by considering the implications for future research.

9.1 Research objectives of this thesis

Before the summary of the chapters, I first restate the three research objectives presented in Chapter 1:

1. Critically review the current comparative effectiveness research which use PPP as IV
2. Implement PPP IV in drug comparison studies using routinely collected data.
3. Explore the novel use of PPP IV in different settings, using real-world data as well as simulated data.

9.2 Summary of thesis chapters

To provide a structure for this discussion, I divided the chapters into three parts:

- Part 1: Introduction and the literature reviews: Chapter 1, Chapter 2, and Chapter 3
- Part 2: CER using real-life studies: Chapter 4, Chapter 5, Chapter 6
- Part 3: CER using simulation studies: Chapter 7, Chapter 8

9.2.1 Part 1

In Chapter 1, I provided a general introduction of IV methods, and specifically, PPP as IVs in CER. I also stated the main objectives of this thesis. In Chapter 2, I firstly summarised the literature focused on implementing IV in comparative effectiveness research, then provide guidance on implementing PPP IV. In Chapter 3, I reviewed the current comparative effectiveness research and simulation studies of which focus is on the using PPP as IV in CER (using real-life data and hypothesised). It covers the research objective 1 and also shed lights on research objective 3 by bring simulation studies into discussion. Both chapters conclude that PPP can be a valid instrumental variable.

9.2.2 Part 2

In part 2, I conducted observational CER using routinely collected health datasets. In Chapter 4, I focused on comparing the effectiveness between diazepam and chlordiazepoxide hydrochloride in preventing AIH rehospitalisation. I did not compare disulfiram and acamprosate as the medications for the alcohol dependence, due to small sample sizes. I found statistical evidence that diazepam is more effective in preventing AIH rehospitalisation than chlordiazepoxide hydrochloride. Due to the limitation of sample size from Chapter 4, I further conducted a CER that compare the disulfiram and acamprosate using a larger study population from CPRD and PIS in Chapter 5. I found evidence in England and Scotland that disulfiram is a better medication than acamprosate in preventing alcohol dependence hospitalisation. In Chapter 6, I conducted a target trial emulation by applying a stricter entrance criterion in the construction of study cohort from a benchmark RCT. Results show that SU is more likely to reduce the HbA1c level to less than 42 mmol/mol in comparison with DPP-4 inhibitor. In Chapter 6, I implemented GP-level prescribing preference as IV because SDRN does not contain the identifier of physicians. Part 2 covers research objective 2 by implementing PPP IV in observational studies. Chapter 4 and Chapter 5 concluded that PPP can be valid IVs in CER. Chapter 6 revealed that the hospital-based prescribing preference can also be a valid IV. However, the IV results in part 2 tend to have lower statistical power compared to the results from conventional approaches.

9.2.3 Part 3

Following the findings from Chapter 3 showing that there is a gap between the simulation studies and applied studies (CER), I conducted simulation studies as an extension for the applied studies. Chapter 7 is inspired by the Chapter 4 where the sample size is moderate (around 2500), and the distribution of treatment assignment is uneven. Chapter 7 revisited research objective 3, with an emphasis on the potential limitation from sample size of the performance of 2SLS. The main finding from Chapter 7 is that 2SLS has potential to reducing the unmeasured confounding bias in nonlinear settings which is consistent with current studies, despite the fact that 2SLS should be used under strict linear assumptions. It also provides evidence that the longer prescription history can better reflect the 'true' prescribing preference of the physicians. I found the proportion of one particular drug prescribed by the physician, which accounts for the 'whole' prescribing history of a physician, tends to be the strongest IV. Chapter 8 is a simulation study that considers both unmeasured confounding bias and noncollapsibility effect in the setting of binary outcome and treatment. Chapter 8 addresses the research objective 3 by accounting for the noncollapsibility effect which has been overlooked in current simulation studies to date on this topic. I explored the performance of 2SRI in the existence of noncollapsibility and found that 2SRI has potential to alleviate the unmeasured confounding bias effect when noncollapsibility effect is not severe. Both Chapter 7 and Chapter 8 are built on the most common scenario in drug comparison studies: binary outcome and binary treatment. Chapter 8 can be viewed as a non-linear extension of Chapter 7.

9.3 Gaps identified in this thesis.

In Table 21, I summarise the evidence gaps I identified in the literature and how I has addressed these.

Inadequate reporting of assumptions in CER IV literature validation of IV	The validation of IV assumptions is often inadequately report in CERs, except for the relevance assumption. Despite the fact that there are multiple approaches to verify the independence assumptions indirectly, there are also studies not reporting it. Reporting on the validation of exclusion restriction assumption is very rare (findings from Chapter 3). This maybe because the prescribing preference of physicians' is believed to be less likely directly associated with the treatment effect on patients. From the findings of Chapter 2, I underscored the necessity of reporting monotonicity assumptions and pointed out the validation of such assumptions are overlooked in most CER using PPP IV.
Gaps between the simulation studies and applied studies	The main finding from Chapter 3 indicates that there is a noticeable methodological gap between the simulation studies and applied studies on this topic. The characteristics of PPP IV lead to extra complexity in the implementation. Therefore, there are many simulation studies focus on exploring the pitfalls of using PPP IV but without empirical examples from real-life studies. Many applied studies tend to simplify the implementation and not consider findings from the simulation studies. I would recommend combining the findings from simulation studies in designing and conducting real-life CER studies. It should also be operated in the other way too – real-life studies informing the simulation studies. For example, the design of the simulation study in Chapter 7 is inspired by the real-life data used in Chapter 4 and in Chapter 8, an empirical case study is presented to verify the findings from simulated studies.
Arbitrary choice of the form of PPP IV	From my findings in Chapter 3, not many studies have attempted different ways of formulating and constructing PPP IVs. They tend to use the most recent prescription or the proportion – these are the 'conventional' ways in the literature. However, my findings from Chapter 3 shows that the strength of IV is strongly influenced by PPP IV formulation. For example, my findings shown in Chapter 4 and Chapter 7 show that different length of

	prescription history tend to be positively associated with the strength of IV; longer prescribing history tends to lead to stronger IVs. Although the numerical PPP IV may bring extra complication to the validation of assumptions, some studies tend to use the numerical PPP IV without presenting validation.
Not enough exploration of PPP IV used in non-linear settings.	Chapter 8 is inspired by the fact that many existing studies ignored the difference between noncollapsibility and confounding where treatment and outcome variables are binary. Whether they implemented 2SRI arbitrarily without considering the non-collapsibility effect may impact the causal inference, or they tend to avoid using IV methods in the non-linear settings due to the complexity. I show that 2SRI is effective in reducing measured confounding but has not been studied extensively from the theoretical perspective.

Table 21. Gaps identified in this thesis.

9.4 Strengths and Limitations

The main strengths of this thesis include: 1) new findings on PPP IV from routinely-collected health data sets; 2) exploration of use of PPP IV in different settings; 3) overview of current CER using PPP IV and providing recommendations for further research in this area; 4) combining IV approach with conventional statistical methods that can only adjust for measured confounding (such as OLS, logistic regression, Cox proportional hazard regression) allowing triangulation between results that do and do not directly address unmeasured confounding. Findings from Chapter 4, Chapter 5 and Chapter 6 provide new high-quality real-world evidence on the comparison of effectiveness of pharmacological treatments. Chapter 7 and Chapter 8 implemented simulation studies based on two scenarios that are rarely investigated on this topic: 1). The performance of 2SLS in reducing unmeasured confounding in moderate and small sample sizes; 2). The ability of 2SRI in alleviating noncollapsibility in non-linear settings. In Chapter 3, I provided a literature review of recent CER that use PPP IV which is the most recent literature review and the only one that also consider the simulation studies so far. I tend to agree with that IV approach can play an important role in the estimating treatment effect. I include IV method as one of statistical approaches to address potential unmeasured confounding as well as other types of statistical methods in Chapter 4, Chapter 5, Chapter 6, Chapter 7, and Chapter 8. The difference between the estimated treatment effect from these approaches can be interpreted as the treatment heterogeneity among the treated and control group and/or potential indirection of unmeasured confounding bias.

One major limitation comes from the IV method itself. Despite the fact that IV methods are proven to be effective in addressing the unmeasured confounding issue in comparative effectiveness research, they can only produce valid estimation when the assumptions hold. However, the validation of assumptions has limitations in that only the relevance assumption can be validated empirically. Besides, PPP IV is a latent variable that require a surrogate or proxy variable; the construction of proxy of PPP IV is highly associated with the performance of the IV methods and estimation of treatment effect. Like other forms of IV, preference based IV comes with the complexity and difficulty in the validation of assumptions.

Another source of limitation is the usual nature of observational studies, in particular the availability of and the composition of data. Data in this thesis comes from routinely-collected datasets. There are cases where the variables in the datasets cannot meet the requirements of constructing PPP, for example, in Chapter 6, where SDRN only includes the identifier for general practice and health boards rather than at the physician level. Further, if data cannot be dated back to a long enough period, it may not capture a prescribing preference to be a valid IV. For some of the analyses in this thesis, this is a cause of missing data.

There were also limitations in terms of setting. The main background of this thesis is CER. The main objective of Chapter 4, Chapter 5, and Chapter 6 is to compare the effectiveness of certain medical treatments. Like many CERs on this topic, IV methods are often used as a supplementary for conventional statistical methods but not a major way to compare the effectiveness. The IV methods conducted in this thesis are mainly two-stage approaches. I did not consider other approaches that can be used for IV methods, such as propensity score calibration, and two-stage calibration statistical approaches. Also, I did not conduct IV methods in CER which compare more than two treatments.

9.5 Implications for further research

Despite the fact that PPP had been shown to be a valid IV in this thesis as well as in other studies, it is still the case in the empirical literature that not many CERs considered using PPP as an IV. I conducted literature research in Google Scholar using key words: ‘physician’s prescribing preference ‘AND ‘instrumental variable’. After 2020, there is no noticeable increase since I conducted the literature review after Chapter 3 (see Figure 18).

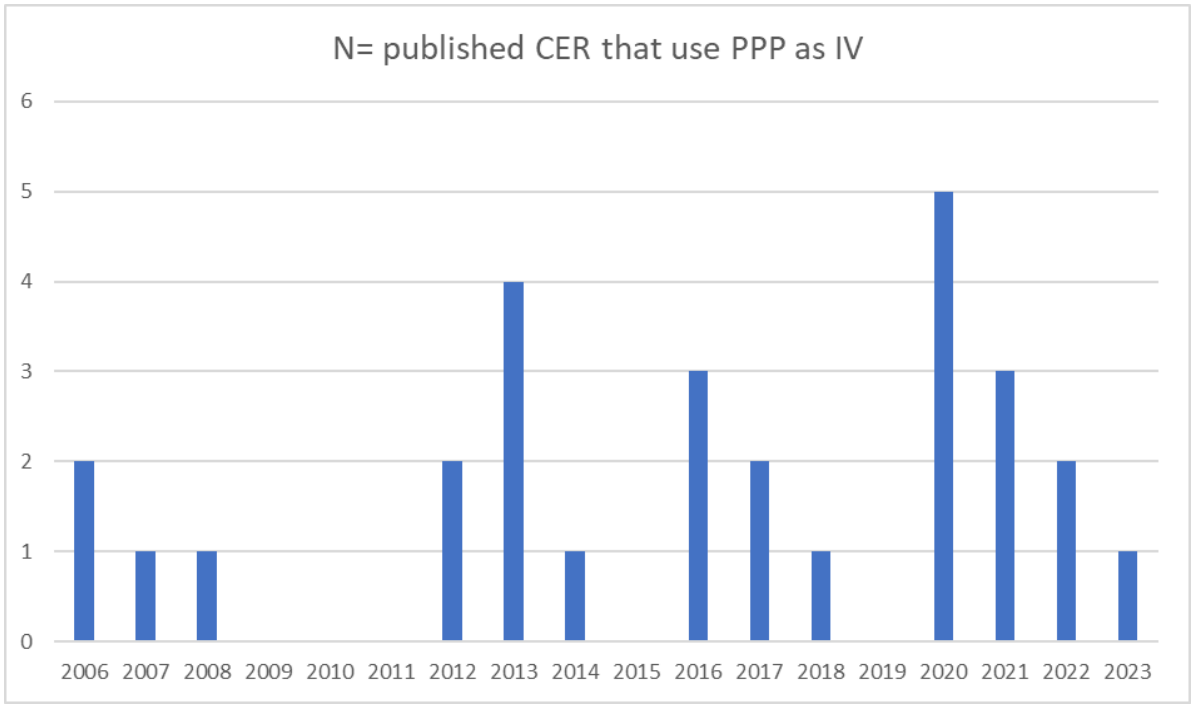


Figure 18. Published CER using PPP as IV from 2005 to 2023

It is important to note that this trend does not cover the CER that use facility-level/hospital-level prescribing preference. In fact, there are multiple CER that implemented other forms of prescribing preference from 2020 to 2023 (for example, hospital-level prescribing preference, centre-level prescribing preference) (Okubo et al., 2021, Littau et al., 2022, Amini et al., 2022, O'Byrne et al., 2023, Wang et al., 2023, Larney et al., 2023). It may be that researchers are cautious when using PPP as IV, and this may be due the limitation of using PPP as IV that had been critically discussed in this chapter and throughout my thesis. The absence of physician level data can be another reason (such as in Chapter 6). This finding is consistent with a recent review on which indicated that there are limited use of IV approach in oncology (Lu et al., 2023).

I agree with conclusions from the current review studies that the PPP IV need to be used with caution. Therefore, I summarise key considerations that one should consider during the implementation of PPP IV in CER in Table 22.

Key considerations	Explanations
Consider the possibility of unmeasured confounding.	The utilisation of IV method should begin with caution and by doing a sensitivity analysis of unmeasured confounding.
Beware of the additional complexity from CER that focus on the medical treatment for rare conditions	CERs that focus on the medical treatments for rare conditions are more likely to involve smaller sample sizes. This will likely cause weaker IVs and weaker statistical power. The physicians may have less prescribing history and it will be difficult to identify a clear prescribing preference. Besides, this may case the large proportion of missing values in the PPP IV as the physicians are less likely to prescribe enough number of prescriptions during a relatively short time, for example six-month or one-year time. For small sample sizes, I recommend pooling data from different countries, or areas to increase the statistical power of the IV results.
IV strength should be a priority in the case of CER with small and moderate sample sizes.	The estimates from strong enough PPP IV are proven consistent when the sample sizes are large. However, when the sample sizes are smaller, for example, in CER of the medications treating rare conditions, the trade-off between variance and bias reduction is more of a consideration as the variance becomes large. In order to maximise the effectiveness of using IV methods, I suggest researchers strive to find stronger PPP IVs in all cases.
The way PPP IVs are constructed (i.e., proxies) allows for flexibility.	One needs to explore different ways of constructing PPP IVs and capture the strongest one as well as the one that satisfy other assumptions.
The importance of reporting validation of IV assumptions	Try to report the validation of all assumptions, including the relevance assumption, exclusion restriction assumption, independence assumption, and monotonicity assumption. Especially for the monotonicity assumption which is often overlooked and strongly associated with the interpretation of IV estimators. Details are presented in Chapter 2.
The importance of triangulating across different approaches	One needs to combine IV methods and conventional approaches in statistical analysis, such as multivariable regression or propensity scores approaches, to provide richer evidence regarding the estimation of treatment effects.

Table 22. Guidance and key considerations for the further research

9.6 Overall conclusion of the thesis

Prescribing preferences estimated from prescribing history can be used as valid instruments in IV CER analyses. I conclude that PPP IV can, with caution applied, be included in observational CERs to complement and potentially enhance non-IV CER approaches.

List of references

ABRAHAMOWICZ, M., BEAUCHAMP, M.-E., IONESCU-ITTU, R., DELANEY, J. A. & PILOTE, L. 2011. Reducing the variance of the prescribing preference-based instrumental variable estimates of the treatment effect. *American journal of epidemiology,* 174**,** 494-502.

ALI, M. S., GROENWOLD, R. H., BELITSER, S. V., PESTMAN, W. R., HOES, A. W., ROES, K. C., DE BOER, A. & KLUNGEL, O. H. 2015. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *Journal of clinical epidemiology,* 68**,** 122-131.

ALVAREZ-MADRAZO, S., MCTAGGART, S., NANGLE, C., NICHOLSON, E. & BENNIE, M. 2016. Data resource profile: the Scottish national prescribing information system (PIS). *International journal of epidemiology,* 45**,** 714.

AMATO, L., MINOZZI, S., VECCHI, S. & DAVOLI, M. 2010. Benzodiazepines for alcohol withdrawal. *Cochrane Database of Systematic Reviews*.

AMINI, M., VAN LEEUWEN, N., EIJKENAAR, F., VAN DE GRAAF, R., SAMUELS, N., VAN OOSTENBRUGGE, R., VAN DEN WIJNGAARD, I. R., VAN DOORMAAL, P. J., ROOS, Y. B. & MAJOIE, C. 2022. Estimation of treatment effects in observational stroke care data: comparison of statistical approaches. *BMC medical research methodology,* 22**,** 103.

ANALYSIS, R. A. 2018. PHE inquiry into the fall in numbers of people in alcohol treatment: findings. *Research and analysis.* Research and analysis.

ANGRIST, J. D. & IMBENS, G. W. 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association,* 90**,** 431-442.

ANGRIST, J. D., IMBENS, G. W. & RUBIN, D. B. 1996. Identification of causal effects using instrumental variables. *Journal of the American statistical Association,* 91**,** 444-455.

ANGRIST, J. D. J. T. E. J. 2004. Treatment effect heterogeneity in theory and practice. 114**,** C52-C83.

ANTONELLI, M., SESTITO, L., TARLI, C. & ADDOLORATO, G. 2022. Perspectives on the pharmacological management of alcohol use disorder: Are the approved medications effective? *European Journal of Internal Medicine*.

ARMSTRONG, K. 2012. Methods in comparative effectiveness research. *Journal of Clinical Oncology,* 30**,** 4208.

AUSTIN, P. C. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research,* 46**,** 399-424.

AUSTIN, P. C. J. S. I. M. 2009. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. 28**,** 3083-3107.

BAHJI, A., BACH, P., DANILEWITZ, M., CROCKFORD, D., EL-GUEBALY, N., DEVOE, D. J. & SAITZ, R. 2022. Comparative efficacy and safety of pharmacotherapies for alcohol withdrawal: a systematic review and network meta-analysis. *Addiction,* 117**,** 2591-2601.

BAIOCCHI, M., CHENG, J. & SMALL, D. S. 2014. Instrumental variable methods for causal inference. *Statistics in medicine,* 33**,** 2297-2340.

BAIOCCHI, M., SMALL, D. S., LORCH, S. & ROSENBAUM, P. R. 2010. Building a stronger instrument in an observational study of perinatal care for premature infants. *journal of the American Statistical Association,* 105**,** 1285-1296.

BAIOCCHI, M., SMALL, D. S., YANG, L., POLSKY, D. & GROENEVELD, P. W. 2012. Near/far matching: a study design approach to instrumental variables. *Health Services Outcomes Research Methodology,* 12**,** 237-253.

BALKE, A. & PEARL, J. 1997. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association,* 92**,** 1171-1176.

BARNES, H. N. & SAMET, J. H. 1997. Brief interventions with substance-abusing patients. *Medical Clinics of North America,* 81**,** 867-879.

BASU, A., COE, N. B. & CHAPMAN, C. G. 2018. 2sls versus 2sri: A ppropriate methods for rare outcomes and/or rare exposures. *Health economics,* 27**,** 937-955.

BIDULKA, P., O'NEILL, S., BASU, A., WILKINSON, S., SILVERWOOD, R. J., CHARLTON, P., BRIGGS, A., ADLER, A. I., KHUNTI, K. & TOMLINSON, L. A. 2021. Protocol for an observational cohort study investigating personalised medicine for intensification of treatment in people with type 2 diabetes mellitus: the PERMIT study. *BMJ open,* 11**,** e046912.

BOEF, A. G., DEKKERS, O. M., VANDENBROUCKE, J. P. & LE CESSIE, S. 2014. Sample size importantly limits the usefulness of instrumental variable methods, depending on instrument strength and level of confounding. *Journal of clinical epidemiology,* 67**,** 1258-1264.

BOEF, A. G., LE CESSIE, S., DEKKERS, O. M., FREY, P., KEARNEY, P. M., KERSE, N., MALLEN, C. D., MCCARTHY, V. J., MOOIJAART, S. P. & MUTH, C. 2016a. Physician's prescribing preference as an instrumental variable. *Epidemiology,* 27**,** 276-283.

BOEF, A. G., SOUVEREIN, P. C., VANDENBROUCKE, J. P., VAN HYLCKAMA VLIEG, A., DE BOER, A., LE CESSIE, S. & DEKKERS, O. M. 2016b. Instrumental variable analysis as a complementary analysis in studies of adverse effects: venous thromboembolism and second-generation versus third-generation oral contraceptives. *Pharmacoepidemiol and drug safety,* 25**,** 317-24.

BOOTH, C. & TANNOCK, I. 2014. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *British journal of cancer,* 110**,** 551-555.

BROOKHART, M. A., RASSEN, J. A. & SCHNEEWEISS, S. 2010. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol and drug Safety,* 19**,** 537-554.

BROOKHART, M. A., RASSEN, J. A., WANG, P. S., DORMUTH, C., MOGUN, H. & SCHNEEWEISS, S. 2007. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? *Medical care***,** S116-S122.

BROOKHART, M. A. & SCHNEEWEISS, S. 2007. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *The international journal of biostatistics,* 3.

BROOKHART, M. A., WANG, P. S., SOLOMON, D. H. & SCHNEEWEISS, S. 2006. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology,* 17**,** 268-75.

BROOKS, J. M. & FANG, G. 2009. Interpreting treatment-effect estimates with heterogeneity and choice: simulation model results. *Clinical therapeutics,* 31**,** 902-919.

CAI, B., SMALL, D. S. & HAVE, T. R. T. 2011. Two-stage instrumental variable methods for estimating the causal odds ratio: analysis of bias. *Statistics in medicine,* 30**,** 1809-1824.

CANIVELL, S., MATA - CASES, M., REAL, J., FRANCH - NADAL, J., VLACHO, B., KHUNTI, K., GRATACÒS, M. & MAURICIO, D. 2019. Glycaemic control after treatment intensification in patients with type 2 diabetes uncontrolled on two or more non-insulin antidiabetic drugs in a real-world setting. *Diabetes, Obesity and Metabolism,* 21**,** 1373-1380.

CHAPMAN, C. G. & BROOKS, J. M. 2016. Treatment effect estimation using nonlinear two-stage instrumental variable estimators: another cautionary note. *Health services research* 51**,** 2375-2394.

CHEN, H., MEHTA, S., APARASU, R., PATEL, A. & OCHOA - PEREZ, M. 2014. Comparative effectiveness of monotherapy with mood stabilizers versus second generation (atypical) antipsychotics for the treatment of bipolar disorder in children and adolescents. *Pharmacoepidemiology and drug safety,* 23**,** 299-308.

CHEN, K., KANG, D., YU, M., ZHANG, R., ZHANG, Y., CHEN, G. & MU, Y. 2018. Direct head-to-head comparison of glycaemic durability of dipeptidyl peptidase - 4 inhibitors and sulphonylureas in patients with type 2 diabetes mellitus: A meta-analysis of long-term randomized controlled trials. *Diabetes, Obesity and Metabolism,* 20**,** 1029-1033.

CHEN, Y. & BRIESACHER, B. A. 2011. Use of instrumental variable in prescription drug research with observational data: a systematic review. *Journal of clinical epidemiology,* 64**,** 687-700.

CHENG, J. & LIN, W. 2018. Understanding causal distributional and subgroup effects with the instrumental propensity score. *American journal of epidemiology,* 187**,** 614-622.

CHISHOLM, J. 1990. The Read clinical classification. *BMJ: British Medical Journal,* 300**,** 1092.

CHUNG, Y.-R., HA, K. H., KIM, H. C., PARK, S. J., LEE, K. & KIM, D. J. 2019. Dipeptidyl peptidase-4 inhibitors versus other antidiabetic drugs added to metformin monotherapy in diabetic retinopathy progression: a real world-based cohort study. *Diabetes and Metabolism,* 43**,** 640-648.

CINELLI, C., FERWERDA , J. & HAZLETT , C. 2021. *sensemakr: Sensitivity Analysis Tools for Regression Models. R package version 0.1.4.* [Online]. Available: https://CRAN.R-project.org/package=sensemakr [Accessed].

CINELLI, C. & HAZLETT, C. 2020. Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B,* 82**,** 39-67.

CLARKE, P. S. & WINDMEIJER, F. J. J. O. T. A. S. A. 2012. Instrumental variable estimators for binary outcomes. 107**,** 1638-1652.

COLLABORATORS, G. A. A. D. U. 2018. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Psychiatry,* 5**,** 987-1012.

D'HOORE, W., BOUCKAERT, A. & TILQUIN, C. 1996. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *Journal of clinical epidemiology,* 49**,** 1429-1433.

DAVIES, N. M., GUNNELL, D., THOMAS, K. H., METCALFE, C., WINDMEIJER, F. & MARTIN, R. M. 2013a. Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. *Journal of clinical epidemiology,* 66**,** 1386-1396.

DAVIES, N. M., SMITH, G. D., WINDMEIJER, F. & MARTIN, R. M. 2013b. Brief Report: Issues in the Reporting and Conduct of Instrumental Variable Studies: A Systematic Review. *Epidemiology***,** 363-369.

DAVIES, N. M., SMITH, G. D., WINDMEIJER, F. & MARTIN, R. M. 2013c. COX-2 selective nonsteroidal anti-inflammatory drugs and risk of gastrointestinal tract complications and myocardial infarction: an instrumental variable analysis. *Epidemiology*, 352-362.

DAVIES, N. M., TAYLOR, A. E., TAYLOR, G. M., ITANI, T., JONES, T., MARTIN, R. M., MUNAFÒ, M. R., WINDMEIJER, F. & THOMAS, K. H. 2020. Varenicline versus nicotine replacement therapy for long-term smoking cessation: an observational study using the Clinical Practice Research Datalink. *Health Technology Assessment (Winchester, England),* 24, 1.

DAVIES, N. M., TAYLOR, G. M., TAYLOR, A. E., JONES, T., MARTIN, R. M., MUNAFÒ, M. R., WINDMEIJER, F. & THOMAS, K. H. 2018. The effects of prescribing varenicline on two-year health outcomes: an observational cohort study using electronic medical records. *Addiction,* 113, 1105-1116.

DAVIES, N. M., THOMAS, K. H., TAYLOR, A. E., TAYLOR, G. M., MARTIN, R. M., MUNAFÒ, M. R. & WINDMEIJER, F. 2017. How to compare instrumental variable and conventional regression analyses using negative controls and bias plots. *International journal of epidemiology,* 46, 2067-2077.

DAY, E. & DALY, C. 2022. Clinical management of the alcohol withdrawal syndrome. *Addiction,* 117, 804-814.

DE SOUSA, A. & DE SOUSA, A. 2005. An open randomized study comparing disulfiram and acamprosate in the treatment of alcohol dependence. *Alcohol and Alcoholism,* 40, 545-8.

DIEHL, A., ULMER, L., MUTSCHLER, J., HERRE, H., KRUMM, B., CROISSANT, B., MANN, K. & KIEFER, F. 2010. Why is disulfiram superior to acamprosate in the routine clinical setting? A retrospective long-term study in 353 alcohol-dependent patients. *Alcohol and Alcoholism,* 45, 271-7.

ENGLAND, P. H. 2021. Estimates of alcohl dependent adults 2018-19. *In:* SHEFFIELD, U. O. (ed.).

ERIKSSON, J. W., BODEGARD, J., NATHANSON, D., THURESSON, M., NYSTRÖM, T. & NORHAMMAR, A. 2016. Sulphonylurea compared to DPP-4 inhibitors in combination with metformin carries increased risk of severe hypoglycemia, cardiovascular events, and all-cause mortality. *Diabetes research clinical practice,* 117, 39-47.

ERTEFAIE, A., FLORY, J. H., HENNESSY, S. & SMALL, D. S. 2017. Instrumental variable methods for continuous outcomes that accommodate nonignorable missing baseline values. *American journal of epidemiology,* 185, 1233-1239.

ESPOSITO, K., COZZOLINO, D., BELLASTELLA, G., MAIORINO, M., CHIODINI, P., CERIELLO, A. & GIUGLIANO, D. 2011. Dipeptidyl peptidase-4 inhibitors and HbA1c target of< 7% in type 2 diabetes: meta - analysis of randomized controlled trials. *Diabetes, Obesity and Metabolism,* 13, 594-603.

EXCELLENCE, N. N. I. F. H. A. C. *Disulfiram* [Online]. Available: https://bnf.nice.org.uk/drugs/disulfiram/ [Accessed].

FADINI, G. P., BOTTIGLIENGO, D., D'ANGELO, F., CAVALOT, F., BOSSI, A. C., ZATTI, G., BALDI, I. & AVOGARO, A. 2018. Comparative effectiveness of DPP-4 inhibitors versus sulfonylurea for the treatment of type 2 diabetes in routine clinical practice: a retrospective multicenter real-world study. *Diabetes Therapy,* 9, 1477-1490.

FANG, G., BROOKS, J. M. & CHRISCHILLES, E. A. 2010. A new method to isolate local-area practice styles in prescription use as the basis for instrumental variables in comparative effectiveness research. *Medical care*, 710-717.

FANG, G., BROOKS, J. M. & CHRISCHILLES, E. A. 2012. Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. *American journal of epidemiology,* 175**,** 60-65.

FARAONI, D. & SCHAEFER, S. T. 2016. Randomized controlled trials vs. observational studies: why not just live together? *BMC anesthesiology,* 16**,** 1-4.

FLURY, B. K. & RIEDWYL, H. 1986. Standard distance in univariate and multivariate analysis. *The American Statistician,* 40**,** 249-251.

FORBES, S. P. & DAHABREH, I. J. 2020. Benchmarking observational analyses against randomized trials: a review of studies assessing propensity score methods. *Journal of general internal medicine,* 35**,** 1396-1404.

FOROUTAN, N., MURATOV, S. & LEVINE, M. 2016. Safety and efficacy of dipeptidyl peptidase-4 inhibitors vs sulfonylurea in metformin-based combination therapy for type 2 diabetes mellitus: systematic review and meta-analysis. *Clinical Investigative Medicine***,** E48-E62.

FRANKLIN, J. M., SCHNEEWEISS, S., HUYBRECHTS, K. F. & GLYNN, R. 2015. Evaluating possible confounding by prescriber in comparative effectiveness research. *Epidemiology,* 26**,** 238-241.

GARABEDIAN, L. F., CHU, P., TOH, S., ZASLAVSKY, A. M. & SOUMERAI, S. B. 2014. Potential bias of instrumental variable analyses for observational comparative effectiveness research. *Annals of internal medicine,* 161**,** 131-138.

GARRIDO, M. M., DEB, P., BURGESS JR, J. F. & PENROD, J. D. 2012. Choosing models for health care cost analyses: issues of nonlinearity and endogeneity. *Health services research,* 47**,** 2377-2397.

GAUTHIER, J., WU, Q. & GOOLEY, T. 2020. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. *Bone marrow transplantation,* 55**,** 675-680.

GLYMOUR, M. M. 2022. Modelling change in a causal framework. *International journal of epidemiology*.

GOMES, M., LATIMER, N., SOARES, M., DIAS, S., BAIO, G., FREEMANTLE, N., DAWOUD, D., WAILOO, A. & GRIEVE, R. 2022. Target trial emulation for transparent and robust estimation of treatment effects for health technology assessment using real-world data: opportunities and challenges. *PharmacoEconomics***,** 1-10.

GOTTLIEB, A., YANOVER, C., CAHAN, A. & GOLDSCHMIDT, Y. 2017. Estimating the effects of second-line therapy for type 2 diabetes mellitus: retrospective cohort study. *BMJ Open Diabetes Research Care,* 5**,** e000435.

GREENACRE, Z. A. 2016. The importance of selection bias in internet surveys. *Open Journal of Statistics,* 6**,** 397.

GREENLAND, S. 1990. Randomization, statistics, and causal inference. *Epidemiology***,** 421-429.

GREENLAND, S. 2000. An introduction to instrumental variables for epidemiologists. *International journal of epidemiology,* 29**,** 722-729.

GREENLAND, S., PEARL, J. & ROBINS, J. M. 1999. Confounding and collapsibility in causal inference. *Statistical science,* 14**,** 29-46.

GUIDELINE, N. 2022. *NICE guideline [NG28]: Type 2 diabetes in adults: management* [Online]. Available: https://www.nice.org.uk/guidance/ng28/chapter/Recommendations#reviewing-drug-treatments [Accessed March 10 2023].

HABER, P., LINTZERIS, N., PROUDE, E. & LOPATKO, O. 2009. Guidelines for the treatment of alcohol problems.

HANEUSE, S. 2016. Distinguishing selection bias and confounding bias in comparative effectiveness research. *Medical care,* 54**,** e23.

HARRIS, K. M. & REMLER, D. K. 1998. Who is the marginal patient? understanding instrumental variables estimates of treatment effects. *Health services research,* 33**,** 1337.

HARRIS, S. K., LOUIS-JACQUES, J. & KNIGHT, J. R. 2014. Screening and brief intervention for alcohol and other abuse. *Adolescent medicine: state of the art reviews,* 25**,** 126-156.

HAUSMAN, J. A. 1978. Specification tests in econometrics. *Econometrica: Journal of the econometric society***,** 1251-1271.

HEALTH, N. C. C. F. M., HEALTH, N. I. F. & EXCELLENCE, C. 2011. Alcohol use disorders: The NICE guideline on the diagnosis, assessment and management of harmful drinking and alcohol dependence.

HECKMAN, J. J. & VYTLACIL, E. 2005. Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica,* 73**,** 669-738.

HECKMAN, J. J. & VYTLACIL, E. J. 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences,* 96**,** 4730-4734.

HERNÁN, M. A. & HERNÁNDEZ-DÍAZ, S. 2012. Beyond the intention-to-treat in comparative effectiveness research. *Clinical Trials,* 9**,** 48-55.

HERNÁN, M. A. & ROBINS, J. M. 2006. Instruments for causal inference: an epidemiologist's dream? *Epidemiology***,** 360-372.

HERNÁN, M. A. & ROBINS, J. M. 2016. Using big data to emulate a target trial when a randomized trial is not available. *American journal of epidemiology,* 183**,** 758-764.

HERNÁN, M. A., SAUER, B. C., HERNÁNDEZ-DÍAZ, S., PLATT, R. & SHRIER, I. 2016. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *Journal of clinical epidemiology,* 79**,** 70-75.

HERRETT, E., GALLAGHER, A. M., BHASKARAN, K., FORBES, H., MATHUR, R., VAN STAA, T. & SMEETH, L. 2015. Data resource profile: clinical practice research datalink (CPRD). *International journal of epidemiology,* 44**,** 827-836.

HOWE, C. J., COLE, S. R., LAU, B., NAPRAVNIK, S. & ERON JR, J. J. 2016. Selection bias due to loss to follow up in cohort studies. *Epidemiology,* 27**,** 91.

HUANG, R., XU, R. & DULAI, P. S. 2020. Sensitivity analysis of treatment effect to unmeasured confounding in observational studies with survival and competing risks outcomes. *Statistics in medicine,* 39**,** 3397-3411.

HUNT, W. A., BARNETT, L. W. & BRANCH, L. G. 1971. Relapse rates in addiction programs. *Journal of Clinical Psychology*.

IMBENS, G. & ANGRIST, J. 1994. Estimation and identification of local average treatment effects. *Econometrica,* 62**,** 467-475.

IONESCU-ITTU, R., ABRAHAMOWICZ, M. & PILOTE, L. 2012. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *Journal of clinical epidemiology,* 65**,** 155-162.

IONESCU-ITTU, R., DELANEY, J. A. & ABRAHAMOWICZ, M. 2009. Bias-variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. *Pharmacoepidemiol and drug safety,* 18**,** 562-71.

JACKSON, J. W. & SWANSON, S. A. 2015. Toward a clearer portrayal of confounding bias in instrumental variable applications. *Epidemiology (Cambridge, Mass.),* 26**,** 498.

JANSSEN, K. J., DONDERS, A. R. T., HARRELL JR, F. E., VERGOUWE, Y., CHEN, Q., GROBBEE, D. E. & MOONS, K. G. 2010. Missing covariate data in medical research: to impute is better than to ignore. *Journal of clinical epidemiology,* 63**,** 721-727.

JAUHAR, P. & ANDERSON, J. 2000. Is daily single dosage of diazepam as effective as chlordiazepoxide in divided doses in alcohol withdrawal—a pilot study. *Alcohol and Alcoholism,* 35**,** 212-214.

JUNG, Y.-C. & NAMKOONG, K. 2014. Alcohol: intoxication and poisoning–diagnosis and treatment. *Handbook of clinical neurology,* 125**,** 115-121.

KANG, H., KREUELS, B., ADJEI, O., KRUMKAMP, R., MAY, J. & SMALL, D. S. 2013. The causal effect of malaria on stunting: a Mendelian randomization and matching approach. *International journal of epidemiology,* 42**,** 1390-1398.

KHUNTI, K., CHARBONNEL, B., COOPER, A., GOMES, M. B., JI, L., LEIGH, P., NICOLUCCI, A., RATHMANN, W., SHESTAKOVA, M. V. & SIDDIQUI, A. 2021. Associations between second-line glucose-lowering combination therapies with metformin and HbA1c, body weight, quality of life, hypoglycaemic events and glucose - lowering treatment intensification: The DISCOVER study. *Diabetes, Obesity and Metabolism,* 23**,** 1823-1833.

KIM, K. J., CHOI, J., LEE, J., BAE, J. H., AN, J. H., KIM, H. Y., YOO, H. J., SEO, J. A., KIM, N. H. & CHOI, K. M. 2019. Dipeptidyl peptidase-4 inhibitor compared with sulfonylurea in combination with metformin: cardiovascular and renal outcomes in a propensity-matched cohort study. *Cardiovascular diabetology,* 18**,** 1-10.

KLEIBER, C. & ZEILEIS, A. 2008. *Applied econometrics with R*, Springer Science & Business Media.

KLUNGEL, O., JAMAL UDDIN, M., DE BOER, A., BELITSER, S., GROENWOLD, R. & ROES, K. 2015. Instrumental variable analysis in epidemiologic studies: an overview of the estimation methods. *Pharm Anal Acta,* 6**,** 2.

KOLADJO, B. F., ESCOLANO, S. & TUBERT-BITTER, P. 2018. Instrumental variable analysis in the context of dichotomous outcome and exposure with a numerical experiment in pharmacoepidemiology. *BMC medical research methodology,* 18**,** 61.

KOLLHORST, B., ABRAHAMOWICZ, M. & PIGEOT, I. 2016. The proportion of all previous patients was a potential instrument for patients' actual prescriptions of nonsteroidal anti-inflammatory drugs. *Journal of clinical epidemiology,* 69**,** 96-106.

KUO, Y.-F., MONTIE, J. E. & SHAHINIAN, V. B. 2012. Reducing bias in the assessment of treatment effectiveness: androgen deprivation therapy for prostate cancer. *Medical care,* 50**,** 374.

LAAKSONEN, E., KOSKI-JÄNNES, A., SALASPURO, M., AHTINEN, H. & ALHO, H. 2008. A randomized, multicentre, open-label, comparative trial of disulfiram, naltrexone and acamprosate in the treatment of alcohol dependence. *Alcohol and Alcoholism,* 43**,** 53-61.

LABRECQUE, J. & SWANSON, S. A. 2018. Understanding the assumptions underlying instrumental variable analyses: a brief review of falsification strategies and related tools. *Current epidemiology reports,* 5**,** 214-220.

LABRECQUE, J. A. & SWANSON, S. A. 2017. Target trial emulation: teaching epidemiology and beyond. *European journal of epidemiology,* 32**,** 473-475.

LARNEY, S., JONES, N. R., HICKMAN, M., NIELSEN, S., ALI, R. & DEGENHARDT, L. 2023. Does opioid agonist treatment reduce overdose mortality risk in people who are older or have

physical comorbidities? Cohort study using linked administrative health data in New South Wales, Australia, 2002–17. *Addiction*.

LEE, J. A., DUBY, J. J. & COCANOUR, C. S. 2019. Effect of early and focused benzodiazepine therapy on length of stay in severe alcohol withdrawal syndrome. *Clinical Toxicology,* 57**,** 624-627.

LEE, K. A., JIN, H. Y., KIM, Y. J., KIM, S. S., CHO, E.-H. & PARK, T. S. 2022. Real-world comparison of mono and dual combination therapies of metformin, sulfonylurea, and dipeptidyl peptidase-4 inhibitors using a common data model: A retrospective observational study. *Medicine,* 101.

LIDDON, L., KINGERLEE, R. & BARRY, J. A. 2018. Gender differences in preferences for psychological treatment, coping strategies, and triggers to help-seeking. *British Journal of Clinical Psychology,* 57**,** 42-58.

LIPSITCH, M., TCHETGEN, E. T. & COHEN, T. 2010. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology (Cambridge, Mass.),* 21**,** 383.

LITTAU, M. J., FREEMAN, R., VIGNESWARAN, W. T., LUCHETTE, F. A., BAKER, M. S., RAAD, W., ABDELSATTAR, Z. M., GRENDA, T., LUBAWSKI, J. & MADARIAGA, M. L. L. 2022. Comparative effectiveness of stereotactic body radiation therapy versus surgery for stage I lung cancer in otherwise healthy patients: an instrumental variable analysis. *JTCVS open,* 9**,** 249-261.

LOUSDAL, M. L. 2018. An introduction to instrumental variable assumptions, validation and estimation. *Emerging themes in epidemiology,* 15**,** 1.

LU, B., THOMSON, S., BLOMMAERT, S., TADROUS, M., EARLE, C. C. & CHAN, K. K. 2023. Use of instrumental variable analyses for evaluating comparative effectiveness in empirical applications of oncology: A systematic review. *Journal of Clinical Oncology,* 41**,** 2362-2371.

MARTENS, E. P., PESTMAN, W. R., DE BOER, A., BELITSER, S. V. & KLUNGEL, O. H. 2006. Instrumental variables: application and limitations. *Epidemiology***,** 260-267.

MARTÍNEZ-CAMBLOR, P., MACKENZIE, T., STAIGER, D. O., GOODNEY, P. P. & O'MALLEY, A. J. 2019. Adjusting for bias introduced by instrumental variable estimation in the Cox proportional hazards model. *Biostatistics,* 20**,** 80-96.

MASON, B. J. & HEYSER, C. J. 2010. Acamprosate: a prototypic neuromodulator in the treatment of alcohol dependence. *CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders),* 9**,** 23-32.

MAYO-SMITH, M. F. 1997. Pharmacological management of alcohol withdrawal: a meta-analysis and evidence-based practice guideline. *Jama,* 278**,** 144-151.

MCGURNAGHAN, S. J., BLACKBOURN, L. A., CAPARROTTA, T. M., MELLOR, J., BARNETT, A., COLLIER, A., SATTAR, N., MCKNIGHT, J., PETRIE, J. & PHILIP, S. 2022. Cohort profile: the Scottish Diabetes Research Network national diabetes cohort–a population-based cohort of people with diabetes in Scotland. *BMJ open,* 12**,** e063046.

MIRIJELLO, A., D'ANGELO, C., FERRULLI, A., VASSALLO, G., ANTONELLI, M., CAPUTO, F., LEGGIO, L., GASBARRINI, A. & ADDOLORATO, G. 2015. Identification and management of alcohol withdrawal syndrome. *Drugs,* 75**,** 353-365.

MOHAN, V., KHUNTI, K., CHAN, S. P., FILHO, F. F., TRAN, N. Q., RAMAIYA, K., JOSHI, S., MITHAL, A., MBAYE, M. N. & NICODEMUS, N. A. 2020. Management of type 2 diabetes in developing countries: balancing optimal glycaemic control and outcomes with affordability and accessibility to treatment. *Diabetes Therapy,* 11**,** 15-35.

MORRIS, T. P., WHITE, I. R. & CROWTHER, M. J. 2019. Using simulation studies to evaluate statistical methods. *Statistics in medicine,* 38**,** 2074-2102.

NALLAMOTHU, B. K., HAYWARD, R. A. & BATES, E. R. 2008. Beyond the randomized clinical trial: the role of effectiveness studies in evaluating cardiovascular therapies. *Circulation,* 118**,** 1294-1303.

NATIONAL COLLABORATING CENTRE FOR MENTAL, H. 2011. National Institute for Health and Care Excellence: Guidelines. *Alcohol-Use Disorders: Diagnosis, Assessment and Management of Harmful Drinking and Alcohol Dependence.* Leicester (UK): British Psychological Society (UK)

NAUCK, M., MEININGER, G., SHENG, D. O., TERRANELLA, L., STEIN, P. P. & GROUP, S. S. 2007. Efficacy and safety of the dipeptidyl peptidase-4 inhibitor, sitagliptin, compared with the sulfonylurea, glipizide, in patients with type 2 diabetes inadequately controlled on metformin alone: a randomized, double-blind, non-inferiority trial. *Diabetes, Obesity and Metabolism,* 9**,** 194-205.

NELSON, R. E., NEBEKER, J. R., HAYDEN, C., REIMER, L., KONE, K. & LAFLEUR, J. 2013. Comparing adherence to two different HIV antiretroviral regimens: an instrumental variable analysis. *AIDS and Behavior,* 17**,** 160-167.

O'BYRNE, M. L., SONG, L., HUANG, J., LEMLEY, B., GOLDBERG, D., GARDNER, M. M., RAVISHANKAR, C., ROME, J. J. & GLATZ, A. C. 2023. Attributable mortality benefit of digoxin treatment in hypoplastic left heart syndrome after the Norwood operation: An instrumental variable-based analysis using data from the Pediatric Health Information Systems Database. *American heart journal,* 263**,** 35-45.

OKUBO, Y., HORIMUKAI, K., MICHIHATA, N., MORITA, K., MATSUI, H., FUSHIMI, K. & YASUNAGA, H. 2021. Association between early antibiotic treatment and clinical outcomes in children hospitalized for asthma exacerbation. *Journal of Allergy Clinical Immunology,* 147**,** 114-122. e14.

ORGANIZATION, W. H. 2014. Global status report on noncommunicable diseases 2014. World Health Organization.

PALMER, T. M., HOLMES, M. V., KEATING, B. J. & SHEEHAN, N. A. 2017. Correcting the standard errors of 2-stage residual inclusion estimators for mendelian randomization studies. *American journal of epidemiology,* 186**,** 1104-1114.

PANG, M., KAUFMAN, J. S. & PLATT, R. W. 2016. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Statistical methods in medical research,* 25**,** 1925-1937.

PATEL, A. K. & BALASANOVA, A. A. 2021. Treatment of Alcohol Use Disorder. *Jama,* 325**,** 596.

POOLE, C., SHRIER, I. & VANDERWEELE, T. J. 2015. Is the risk difference really a more heterogeneous measure? *Epidemiology,* 26**,** 714-718.

POTTER, B. J., DORMUTH, C. & LE LORIER, J. 2020. A theoretical exploration of therapeutic monomania as a physician-based instrumental variable. *Pharmacoepidemiol and drug safety* 29**,** 45-52.

PRATT, N., ROUGHEAD, E. E., RYAN, P. & SALTER, A. 2010. Antipsychotics and the risk of death in the elderly: an instrumental variable analysis using two preference based instruments. *Pharmacoepidemiol and drug safety,* 19**,** 699-707.

RASSEN, J. A., BROOKHART, M. A., GLYNN, R. J., MITTLEMAN, M. A. & SCHNEEWEISS, S. 2009a. Instrumental variables I: instrumental variables exploit natural variation in

nonexperimental data to estimate causal relationships. *Journal of clinical epidemiology,* 62**,** 1226-1232.

RASSEN, J. A., SCHNEEWEISS, S., GLYNN, R. J., MITTLEMAN, M. A. & BROOKHART, M. A. 2009b. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *American journal of epidemiology,* 169**,** 273-284.

REID, M. C., FIELLIN, D. A. & O'CONNOR, P. G. 1999. Hazardous and harmful alcohol consumption in primary care. *Archives of internal medicine,* 159**,** 1681-1689.

ROBINS, J. M. 1994. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory methods,* 23**,** 2379-2412.

ROSENBAUM, P. R. & RUBIN, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70**,** 41-55.

RUAN, Z., ZOU, H., LEI, Q., UNG, C. O. L., SHI, H. & HU, H. 2022. Pharmacoeconomic evaluation of dipeptidyl peptidase-4 inhibitors for the treatment of type 2 diabetes mellitus: a systematic literature review. *Expert Review of Pharmacoeconomics Outcomes Research*

1-20.

RUBIN, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology,* 66**,** 688.

RUBIN, D. B. 1986. Comment: Which ifs have causal answers. *Journal of the American statistical association,* 81**,** 961-962.

SCHEEN, A. 2020. Reduction in HbA1c with SGLT2 inhibitors vs. DPP-4 inhibitors as add-ons to metformin monotherapy according to baseline HbA1c: A systematic review of randomized controlled trials. *Diabetes Metabolism,* 46**,** 186-196.

SCHMIDT, K. J., DOSHI, M. R., HOLZHAUSEN, J. M., NATAVIO, A., CADIZ, M. & WINEGARDNER, J. E. 2016. Treatment of severe alcohol withdrawal. *Annals of Pharmacotherapy,* 50**,** 389-401.

SCHNEEWEISS, S. 2007. Developments in post-marketing comparative effectiveness research. *Clin Pharmacol Ther,* 82**,** 143-56.

SCHNEEWEISS, S., SEEGER, J. D., LANDON, J. & WALKER, A. M. 2008. Aprotinin during coronary-artery bypass grafting and risk of death. *New England Journal of Medicine,* 358**,** 771-83.

SCHNEEWEISS, S., SETOGUCHI, S., BROOKHART, A., DORMUTH, C. & WANG, P. S. 2007. Risk of death associated with the use of conventional versus atypical antipsychotic drugs among elderly patients. *Cmaj,* 176**,** 627-632.

SCHNEEWEISS, S., SOLOMON, D. H., WANG, P. S., RASSEN, J. & BROOKHART, M. A. 2006. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective cyclooxygenase 2 inhibitors and nonselective nonsteroidal antiinflammatory drugs: an instrumental variable analysis. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology,* 54**,** 3390-3398.

SCHUSTER, N. A., TWISK, J. W., TER RIET, G., HEYMANS, M. W. & RIJNHART, J. J. 2021. Noncollapsibility and its role in quantifying confounding bias in logistic regression. *BMC medical research methodology,* 21**,** 1-9.

SCHWARTZ, S., GATTO, N. M. & CAMPBELL, U. B. 2012. Extending the sufficient component cause model to describe the Stable Unit Treatment Value Assumption (SUTVA). *Epidemiologic Perspectives & Innovations,* 9**,** 1-11.

SCOTLAND, N. R. O. Statistics and Data.

SCOTLAND, P. H. 2021. *Alcohol-Related Hospital*

Statistics Scotland 2020/21 [Online]. Available: https://publichealthscotland.scot/publications/alcohol-related-hospital-statistics/alcohol-related-hospital-statistics-scotland-financial-year-2020-to-2021/ [Accessed 2022].

SECEMSKY, E. A., KIRTANE, A., BANGALORE, S., JOVIN, I. S., PATEL, D., FERRO, E. G., WIMMER, N. J., ROE, M., DAI, D. & MAURI, L. 2017. Practice patterns and in-hospital outcomes associated with bivalirudin use among patients with non–ST-segment–elevation myocardial infarction undergoing percutaneous coronary intervention in the United States. *Circulation: Cardiovascular Quality and Outcomes,* 10**,** e003741.

SHAH, B. R., DROZDA, J. & PETERSON, E. D. 2010. Leveraging observational registries to inform comparative effectiveness research. *American heart journal,* 160**,** 8-15.

SMALL, D. S., TAN, Z., RAMSAHAI, R. R., LORCH, S. A. & BROOKHART, M. A. 2017. Instrumental variable estimation with a stochastic monotonicity assumption. *Statistical science,* 32**,** 561-579.

SØRENSEN, H. T., LASH, T. L. & ROTHMAN, K. J. 2006. Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. *Hepatology,* 44**,** 1075-1082.

STAIGER, D. O. & STOCK, J. H. 1994. Instrumental variables regression with weak instruments. National Bureau of Economic Research Cambridge, Mass., USA.

STATISTICS, O. F. N. 2021. *Alcohol-specific deaths in the UK: registered in 2020* [Online]. Office for National Statistics. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/bulletins/alcoholrelateddeathsintheunitedkingdom/registeredin2020 [Accessed March 15 2022].

STATISTICS, O. F. N. 2022. *Alcohol-specific deaths in the UK: registered iin 2021* [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/alcoholspecificdeathsintheuk/2021registrations [Accessed].

STIRBU, I., KUNST, A. E., MIELCK, A. & MACKENBACH, J. P. 2011. Inequalities in utilisation of general practitioner and specialist services in 9 European countries. *BMC health services research,* 11**,** 1-8.

STOCK, J. H., WRIGHT, J. H. & YOGO, M. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business Economic Statistics,* 20**,** 518-529.

STOCK, J. H. & YOGO, M. 2002. Testing for weak instruments in linear IV regression. National Bureau of Economic Research.

STUART, E. A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics,* 25**,** 1.

SUH, H. S., HAY, J. W., JOHNSON, K. A. & DOCTOR, J. N. 2012. Comparative effectiveness of statin plus fibrate combination therapy and statin monotherapy in patients with type 2 diabetes: use of propensity-score and instrumental variable methods to adjust for treatment-selection bias. *Pharmacoepidemiol and drug safety,* 21**,** 470-84.

SWANSON, S. A. & HERNÁN, M. A. 2013. Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology,* 24**,** 370-374.

SWANSON, S. A. & HERNÁN, M. A. 2014. Think globally, act globally: an epidemiologist's perspective on instrumental variable estimation. *Statistical science: a review journal of the Institute of Mathematical Statistics,* 29**,** 371.

SWANSON, S. A. & HERNÁN, M. A. 2018. The challenging interpretation of instrumental variable estimates under monotonicity. *International journal of epidemiology,* 47**,** 1289-1297.

SWANSON, S. A., MILLER, M., ROBINS, J. M. & HERNÁN, M. A. 2015a. Definition and evaluation of the monotonicity condition for preference-based instruments. *Epidemiology (Cambridge, Mass.),* 26**,** 414.

SWANSON, S. A., ROBINS, J. M., MILLER, M. & HERNÁN, M. A. 2015b. Selecting on treatment: a pervasive form of bias in instrumental variable analyses. *American journal of epidemiology,* 181**,** 191-197.

TAN, X., YANG, L., KHUNTI, K., ZHANG, R., ZHANG, Y., RAJPATHAK, S. & YU, M. 2021. Factors associated with switching from sulphonylureas to dipeptidyl peptidase 4 inhibitors among patients with type 2 diabetes in the United States. *Diabetes, Obesity and Metabolism,* 23**,** 2251-2260.

TAN, Z. 2006. Regression and weighting methods for causal inference using instrumental variables. *Journal of the American Statistical Association,* 101**,** 1607-1618.

TAYLOR, G. M., TAYLOR, A. E., THOMAS, K. H., JONES, T., MARTIN, R. M., MUNAFO, M. R., WINDMEIJER, F. & DAVIES, N. M. 2017. The effectiveness of varenicline versus nicotine replacement therapy on long-term smoking cessation in primary care: a prospective cohort study of electronic medical records. *International journal of epidemiology,* 46**,** 1948-1957.

TCHETGEN, E. J. T., WALTER, S., VANSTEELANDT, S., MARTINUSSEN, T. & GLYMOUR, M. 2015. Instrumental variable estimation in a survival context. *Epidemiology (Cambridge, Mass.),* 26**,** 402.

TENNANT, P. W., ARNOLD, K. F., ELLISON, G. & GILTHORPE, M. S. 2021. Analyses of 'change scores' do not estimate causal effects in observational data. *International journal of epidemiology*.

TERZA, J. V. 2018. Two-stage residual inclusion estimation in health services research and health economics. *Health services research,* 53**,** 1890-1899.

TERZA, J. V., BASU, A. & RATHOUZ, P. J. 2008a. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of health economics,* 27**,** 531-543.

TERZA, J. V., BRADFORD, W. D. & DISMUKE, C. E. 2008b. The use of linear instrumental variables methods in health services research and health economics: a cautionary note. *Health services research,* 43**,** 1102-1120.

THOMPSON, A., WRIGHT, A. K., ASHCROFT, D. M., VAN STAA, T. P. & PIRMOHAMED, M. 2017. Epidemiology of alcohol dependence in UK primary care: Results from a large observational study using the Clinical Practice Research Datalink. *PLoS One,* 12**,** e0174818.

THOMPSON, C. A. & ARAH, O. A. 2014. Selection bias modeling using observed data augmented with imputed record-level probabilities. *Annals of epidemiology,* 24**,** 747-753.

TIERNEY, N. J., HARDEN, F. A., HARDEN, M. J. & MENGERSEN, K. L. 2015. Using decision trees to understand structure in missing data. *BMJ open,* 5**,** e007450.

TULLOCH, J. S., BEADSWORTH, M. B., VIVANCOS, R., RADFORD, A. D., WARNER, J. C. & CHRISTLEY, R. M. 2020. GP coding behaviour for non-specific clinical presentations: a pilot study. *BJGP open,* 4.

UDDIN, M. J., GROENWOLD, R. H., DE BOER, A., AFONSO, A. S., PRIMATESTA, P., BECKER, C., BELITSER, S. V., HOES, A. W., ROES, K. C. & KLUNGEL, O. H. 2016a. Evaluating different physician's prescribing preference based instrumental variables in two primary care

databases: a study of inhaled long-acting beta2-agonist use and the risk of myocardial infarction. *pharmacoepidemiology and drug safety,* 25**,** 132-141.

UDDIN, M. J., GROENWOLD, R. H., DE BOER, A., BELITSER, S. V., ROES, K. C., HOES, A. W. & KLUNGEL, O. H. 2014. Performance of instrumental variable methods in cohort and nested case–control studies: a simulation study. *Pharmacoepidemiology and drug safety,* 23**,** 165-177.

UDDIN, M. J., GROENWOLD, R. H., DE BOER, A., GARDARSDOTTIR, H., MARTIN, E., CANDORE, G., BELITSER, S. V., HOES, A. W., ROES, K. C. & KLUNGEL, O. H. 2016b. Instrumental variables analysis using multiple databases: an example of antidepressant use and risk of hip fracture. *Pharmacoepidemiol and drug safety,* 25 Suppl 1**,** 122-31.

VAN BUUREN, S. & GROOTHUIS-OUDSHOORN, K. 2011. mice: Multivariate imputation by chained equations in R. *Journal of statistical software,* 45**,** 1-67.

VANAGAS, G. 2004. Receiver operating characteristic curves and comparison of cardiac surgery risk stratification systems. *Interactive cardiovascular thoracic surgery,* 3**,** 319-322.

VANDYKE, R. D., MCPHAIL, G. L., HUANG, B., FENCHEL, M. C., AMIN, R. S., CARLE, A. C., CHINI, B. A. & SEID, M. 2013. Inhaled tobramycin effectively reduces FEV1 decline in cystic fibrosis. An instrumental variables analysis. *Annals of the American Thoracic Society,* 10**,** 205-212.

VONGHIA, L., LEGGIO, L., FERRULLI, A., BERTINI, M., GASBARRINI, G., ADDOLORATO, G. & GROUP, A. T. S. 2008. Acute alcohol intoxication. *European Journal of Internal Medicine,* 19**,** 561-567.

WALD, A. J. T. A. O. M. S. 1940. The fitting of straight lines if both variables are subject to error. 11**,** 284-300.

WALKER, V. M., DAVIES, N. M., MARTIN, R. M. & KEHOE, P. G. 2020. Comparison of antihypertensive drug classes for dementia prevention. *Epidemiology (Cambridge, Mass.),* 31**,** 852.

WALTON, M. A., CHERMACK, S. T., SHOPE, J. T., BINGHAM, C. R., ZIMMERMAN, M. A., BLOW, F. C. & CUNNINGHAM, R. M. 2010. Effects of a brief intervention for reducing violence and alcohol misuse among adolescents: a randomized controlled trial. *Jama,* 304**,** 527-535.

WAN, F., SMALL, D. & MITRA, N. 2018. A general approach to evaluating the bias of 2-stage instrumental variable estimators. *Statistics in medicine,* 37**,** 1997-2015.

WANG, A., NIANOGO, R. A. & ARAH, O. A. 2017a. G-computation of average treatment effects on the treated and the untreated. *BMC medical research methodology*

17**,** 1-5.

WANG, H., CORDINER, R. L., HUANG, Y., DONNELLY, L., HAPCA, S., COLLIER, A., MCKNIGHT, J., KENNON, B., GIBB, F. & MCKEIGUE, P. 2023. Cardiovascular Safety in Type 2 Diabetes With Sulfonylureas as Second-Line Drugs: A Nation-Wide Population-Based Comparative Safety Study. *Diabetes Care***,** dc221238.

WANG, J., WU, H.-Y. & CHIEN, K.-L. 2022. Cardioprotective effects of dipeptidyl peptidase-4 inhibitors versus sulfonylureas in addition to metformin: A nationwide cohort study of patients with type 2 diabetes. *Diabetes Metabolism*

48**,** 101299.

WANG, L., ROBINS, J. M. & RICHARDSON, T. S. 2017b. On falsification of the binary instrumental variable model. *Biometrika,* 104**,** 229-236.

WANG, Y., HUNT, K., NAZARETH, I., FREEMANTLE, N. & PETERSEN, I. 2013. Do men consult less than women? An analysis of routinely collected UK general practice data. *BMJ open,* 3**,** e003320.

WEINTRAUB, S. 2017. Diazepam in the treatment of moderate to severe alcohol withdrawal. *CNS drugs,* 31**,** 87-95.

WHO. 2022. *Alcohol* [Online]. [Accessed 18 August 2022].

WIDDING-HAVNERAAS, T. & ZACHRISSON, H. D. 2022. A Gentle Introduction to Instrumental Variables. *Journal of clinical epidemiology*.

WILDING, J., GODEC, T., KHUNTI, K., POCOCK, S., FOX, R., SMEETH, L., CLAUSON, P., FENICI, P., HAMMAR, N. & MEDINA, J. 2018. Changes in HbA1c and weight, and treatment persistence, over the 18 months following initiation of second-line therapy in patients with type 2 diabetes: results from the United Kingdom Clinical Practice Research Datalink. *BMC medicine,* 16**,** 1-12.

WOLF, A., DEDMAN, D., CAMPBELL, J., BOOTH, H., LUNN, D., CHAPMAN, J. & MYLES, P. 2019. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *International journal of epidemiology,* 48**,** 1740-1740g.

WOOLDRIDGE, J. M. 2010. *Econometric analysis of cross section and panel data*, MIT press.

ZHANG, L., LEWSEY, J. & MCALLISTER, D. A. 2022. Comparative effectiveness research considered methodological insights from simulation studies in physician's prescribing preference. *Journal of clinical epidemiology,* 148**,** 74-80.

ZHANG, Z., UDDIN, M. J., CHENG, J. & HUANG, T. 2018. Instrumental variable analysis in the presence of unmeasured confounding. *Annals of translational medicine,* 6.

Supplementary material

List of tables in supplementary material

Table S1. Summarises of the 6 simulation studies that were identified ····· 178

Table S2. Summaries of the CERs that were identified ····· 178

Table S3. Validation of monotonicity assumption ····· 180

Table S4. Values of the IV estimands ····· 182

Table S5. Inequality models on prescriptions before first AUD hospitalization ····· 188

Table S6. Spline age graphs ····· 189

Table S7. Validation of IV assumptions. IVs are dichotomised at mean value ····· 191

Table S8. Read codes included for defining alcohol dependence ····· 202

Table S9. Validation of monotonicity assumption ····· 210

List of figures in supplementary material

Figure S1. SMD based on the treatment ····· 179

Figure S2. SMD based on the dichotomised IV (prior 20) ····· 179

Figure S3. Sensitivity analysis to unmeasured confounding. Outcome is the AIH rehospitalization ····· 181

Figure S4. sensitivity analysis to unmeasured confounding. Outcome is the AIH death ···· 181

Figure S5. Trends in percentage of prescriptions by sex (primary care) ····· 186

Figure S6. Trends in percentage of prescriptions by sex (secondary care) ····· 186

Figure S7. Trends in percentage of prescriptions by age (primary care) ····· 187

Figure S8. Trends in percentage of prescriptions by age (secondary care) ····· 187

Figure S9. Sensitivity analysis to the unmeasured confounding (Chapter 5) ····· 191

Figure S10. SMD of covariates based on the level of treatment (left), SMD of covariates based on the level of IV (right). The dash line represents represent SMD equals 0.1. ····· 210

Figure S11. sensitivity analysis to unmeasured confounding (Chapter 6) ····· 211

Supplementary material for Chapter 3

Two tables can be found on https://www.jclinepi.com/article/S0895-4356(22)00103-2/fulltext#supplementaryMaterial.

Table S1. Summarises of the 6 simulation studies that were identified.

Table S2. Summaries of the CERs that were identified.

Supplementary material for Chapter 4

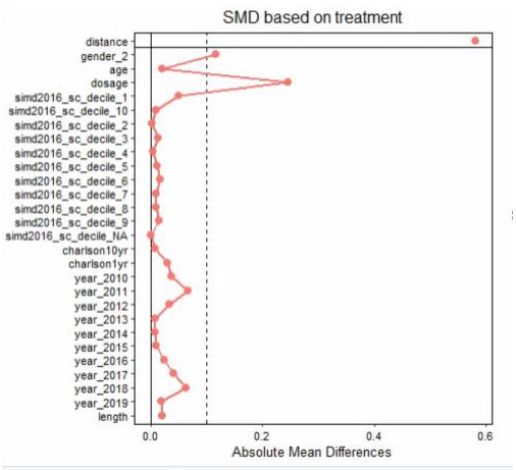


Figure S1. SMD based on the treatment.

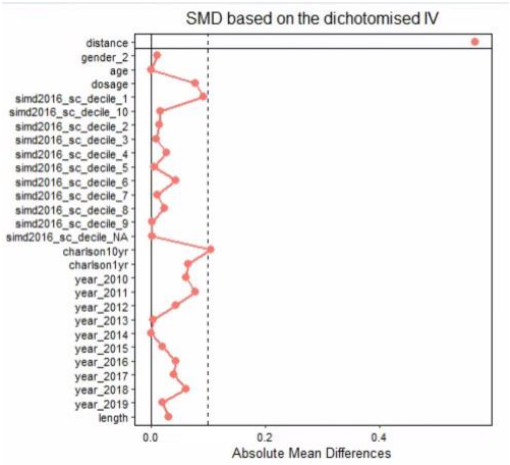


Figure S2. SMD based on the dichotomised IV (prior 20)

IV	Outcome: AIH rehospitalisation			Outcome: AIH death		
IV	IV inequality hold	Monotonicity inequality hold	Wu-Hausman test (p value)	IV inequality hold	Monotonicity inequality hold	Wu-Hausman test (p value)
Prior 1	TRUE	FALSE	0.034	TRUE	FALSE	0.39
Prior 2	TRUE	FALSE	0.016	TRUE	FALSE	0.048
Prior 3	TRUE	FALSE	0.122	TRUE	FALSE	0.159
Prior 4	TRUE	TRUE	0.695	TRUE	FALSE	0.196
Prior 5	TRUE	TRUE	0.383	TRUE	FALSE	0.221
Prior 6	TRUE	TRUE	0.471	TRUE	FALSE	0.018
Prior 7	TRUE	TRUE	0.730	TRUE	FALSE	0.041
Prior 8	TRUE	TRUE	0.526	TRUE	FALSE	0.106
Prior 9	TRUE	TRUE	0.801	TRUE	FALSE	0.060
Prior 10	TRUE	TRUE	0.764	TRUE	FALSE	0.106
Prior 11	TRUE	TRUE	0.806	TRUE	FALSE	0.061
Prior 12	TRUE	TRUE	0.986	TRUE	FALSE	0.049
Prior 13	TRUE	TRUE	0.640	TRUE	FALSE	0.055
Prior 14	TRUE	TRUE	0.774	TRUE	FALSE	0.032
Prior 15	TRUE	TRUE	0.758	TRUE	FALSE	0.030
Prior 16	TRUE	TRUE	0.801	TRUE	FALSE	0.025
Prior 17	TRUE	TRUE	0.745	TRUE	FALSE	0.019
Prior 18	TRUE	TRUE	0.577	TRUE	FALSE	0.016
Prior 19	TRUE	TRUE	0.566	TRUE	FALSE	0.019
Prior 20	TRUE	TRUE	0.688	TRUE	FALSE	0.028
Proportion IV	TRUE	TRUE	0.002	TRUE	TRUE	0.066

Table S3. Validation of monotonicity assumption

```
Sensitivity Analysis to Unobserved Confounding

Model Formula: rehosp ~ drug + age + gender + dosage + year + charlson10yr +
    charlson1yr + length + simd2016_sc_decile

Null hypothesis: q = 1 and reduce = TRUE

Unadjusted Estimates of ' drug ':
  Coef. estimate: -0.03431
  Standard Error: 0.01308
  t-value: -2.62297

Sensitivity Statistics:
  Partial R2 of treatment with outcome: 0.0024
  Robustness Value, q = 1 : 0.04784
  Robustness Value, q = 1 alpha = 0.05 : 0.01229
```

Figure S3. Sensitivity analysis to unmeasured confounding. Outcome is the AIH rehospitalisation. Result from the R function 'sensemarkr'. More detailed explanation of the results can be found in https://cran.r-project.org/web/packages/sensemakr/vignettes/sensemakr.html

```
Sensitivity Analysis to Unobserved Confounding

Model Formula: death ~ drug + age + gender + dosage + year + charlson10yr +
    charlson1yr + length + simd2016_sc_decile

Null hypothesis: q = 1 and reduce = TRUE
-- This means we are considering biases that reduce the absolute value of the current estimate.
-- The null hypothesis deemed problematic is H0:tau = 0

Unadjusted Estimates of 'drug':
  Coef. estimate: -0.0238
  Standard Error: 0.0096
  t-value (H0:tau = 0): -2.4828

Sensitivity Statistics:
  Partial R2 of treatment with outcome: 0.0021
  Robustness Value, q = 1: 0.0453
  Robustness Value, q = 1, alpha = 0.05: 0.0097
```

Figure S4. sensitivity analysis to unmeasured confounding. Outcome is the AIH death. Results from the R function 'sensemakr'.

IV	Non-dichotomised PPP IV	Dichotomised PPP IV	Non-dichotomised PPP IV	Dichotomised PPP IV
	AIH rehospitalisation as outcome		AIH Death as outcome	
Prior 1	-0.922	-0.922	-0.170	-0.170
Prior 2	-0.558	-0.567	-0.203	-0.220
Prior 3	-0.547	-0.426	-0.161	-0.166
Prior 4	-0.405	-0.337	-0.134	-0.139
Prior 5	-0.445	-0.373	-0.121	-0.126
Prior 6	-0.452	-0.351	-0.183	-0.166
Prior 7	-0.447	-0.328	-0.154	-0.136
Prior 8	-0.455	-0.367	-0.123	-0.114
Prior 9	-0.467	-0.350	-0.138	-0.094
Prior 10	-0.469	-0.379	-0.122	-0.0796
Prior 11	-0.504	-0.363	-0.135	-0.101
Prior 12	-0.491	-0.331	-0.135	-0.105
Prior 13	-0.490	-0.324	-0.128	-0.101
Prior 14	-0.447	-0.300	-0.134	-0.110
Prior 15	-0.434	-0.288	-0.134	-0.102
Prior 16	-0.425	-0.274	-0.134	-0.109
Prior 17	-0.421	-0.263	-0.138	-0.0994
Prior 18	-0.416	-0.270	-0.139	-0.104
Prior 19	-0.421	-0.296	-0.133	-0.111
Prior 20	-0.430	-0.284	-0.129	-0.0985
Proportion IV	-0.332	-0.261	-0.112	-0.059

Table S4. Values of the IV estimands

Data request form for Chapter 4

Public Benefit and Privacy Panel for Health and Social Care

PBPP Amendment Request Form – *to be completed for changes to an approved PBPP application*

Application Control - *please complete all sections below*			
Application Coordinator	Johanna Bruce		
Application Number	1718-0238	Approval date	
Applicant name	Prof James Lewsey		
Applicant email address	James.lewsey@glasgow.ac.uk		
Proposal name	TRends and Inequalities in Prescribing for Alcohol Dependence in Scotland (TRIPADS)		
Amendment submission date	01/10/2019		
Summary of amendment (including justification /explanation of changes)	*We request that a PhD student, Lisong Zhang, who is working on a key methodology is added to be able to access the TRIPADS data sets. Lisong will collaborate with us to ensure that the last research question of TRIPADS is answered using cutting-edge techniques.* *We also request that Jill Pell is removed access as she will not be analysing the data.*		
Supporting Documents (please ensure that the original approval letter is attached with all submissions)	☒ Original Approval letter ☒ Updated application with version number ☐ IG certificates (if applicable) ☐ Other (please detail)		

Please tick below to indicate which sections of the application have been amended and note that all changes should be highlighted within the form and an indication of the change noted below:	
x☐ Section 1 (People)	Lisong Zhang to be added MRC Medical Research Council Regulatory Support Centre This is to certify that: **Lisong Zhang** **Passed** **Research, GDPR and confidentiality Quiz** Date / Time: October 24, 2019 3:40 pm Student Score: 88.88 Passing Score: 70 Result: Pass Jill Pell to be removed

☐ Section 2 (Organisations)	
☐ Section 3 (Overview)	
☐ Section 4 (Data and Data Subjects)	
☐ Section 5 (Data Processing)	

Please signify below confirming that other than the changes requested all other information on the original application has not changed.

To be signified by the APPLICANT

Name (in Capitals): JAMES LEWSEY	Date: 01/10/2019

Supplementary material for Chapter 5

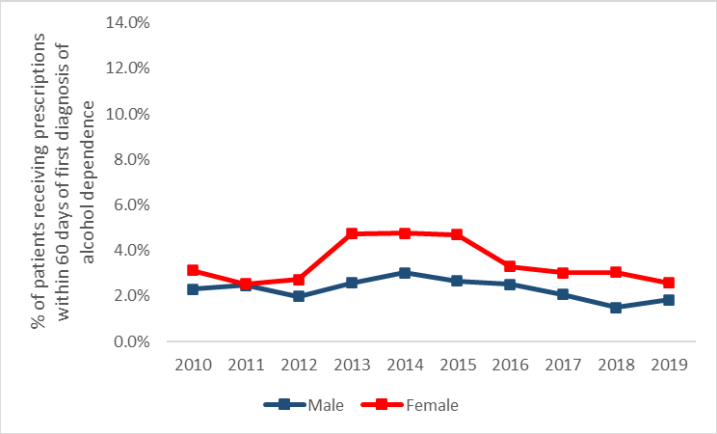


Figure S5. Trends in percentage of prescriptions by sex (primary care)

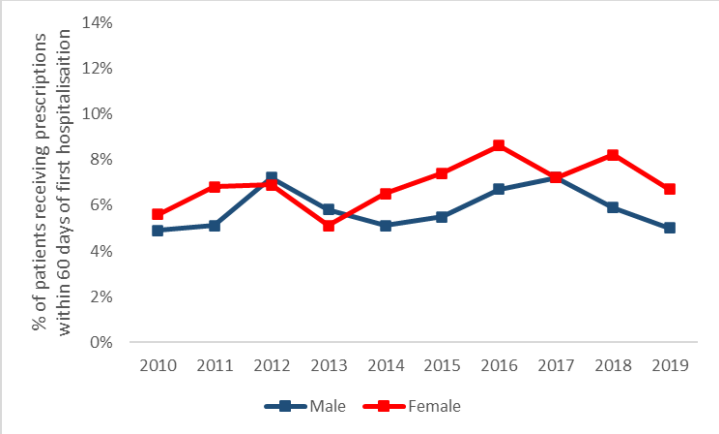


Figure S6. Trends in percentage of prescriptions by sex (secondary care)

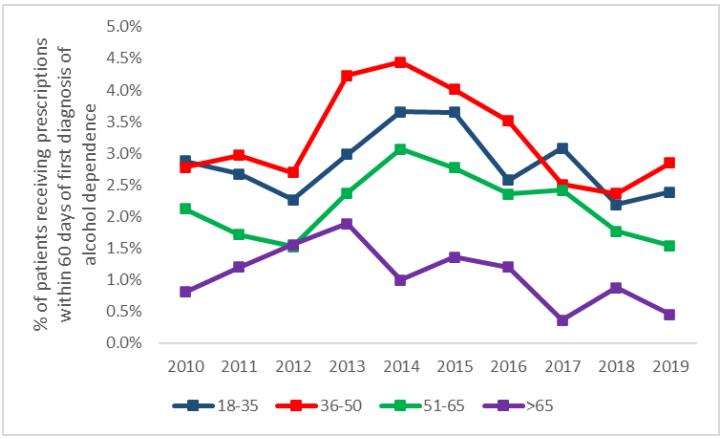


Figure S7. Trends in percentage of prescriptions by age (primary care)

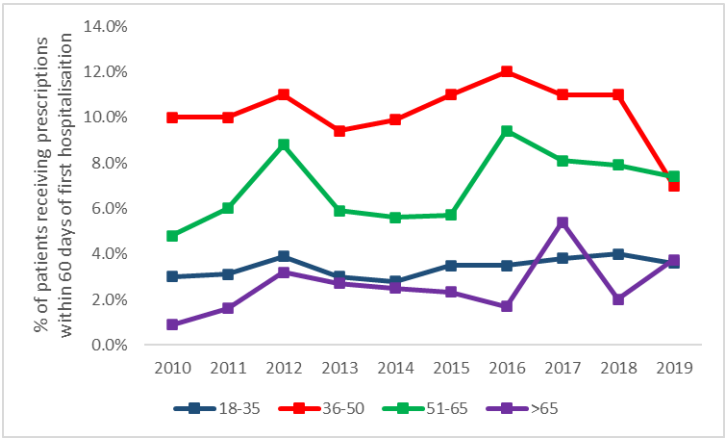
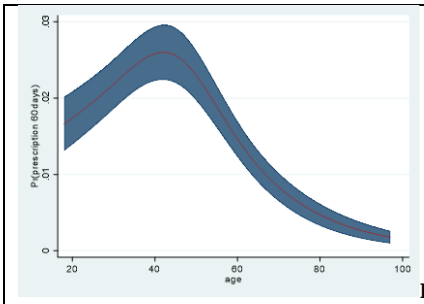


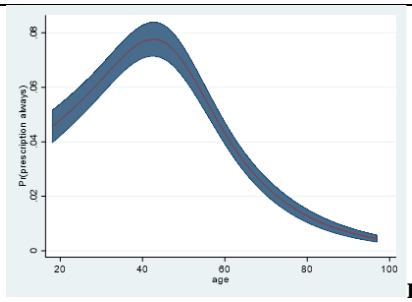
Figure S8. Trends in percentage of prescriptions by age (secondary care)

	Any alcohol dependence prescriptions 60 days prior first hospitalisation	Any alcohol dependence prescription ever before first hospitalisation
Sex (female reference)	0.794 (.673-.937)	0.716 (.654-.783)
Charlson Comorbidity Index (0 reference)		
>=1	1.039 (.870-1.242)	1.092 (.991-1.202)
Mental health comorbidity	1.281 (1.085-1.512)	1.317 (1.203-1.442)
socioeconomic deprivation (1-most deprived- reference)		
2	1.565 (1.254-1.953)	1.275 (1.132-1.437)
3	1.329 (1.040-1.698)	1.318 (1.160-1.498)
4	1.517 (1.172-1.964)	1.372 (1.194-1.576)
5	1.628 (1.229-2.157)	1.502 (1.290-1.750)

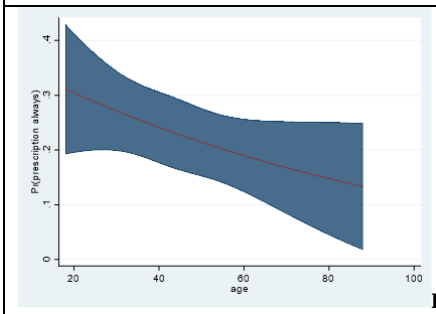
Table S5. Inequality models on prescriptions before first AUD hospitalisation



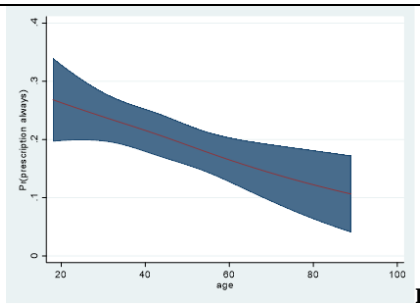
model 1



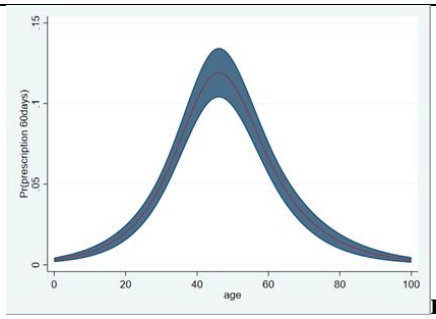
model 2



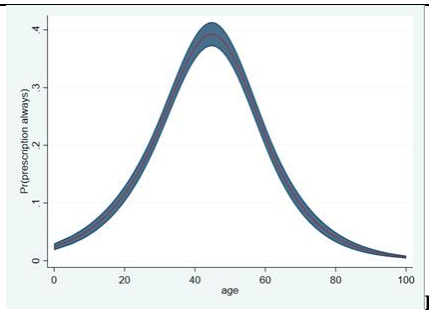
model 3



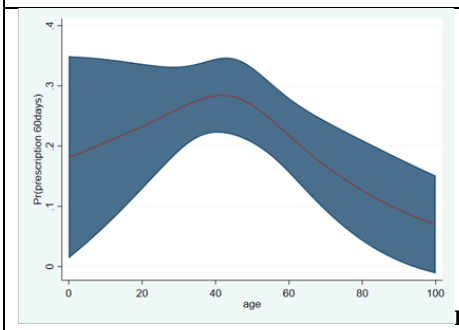
model 4



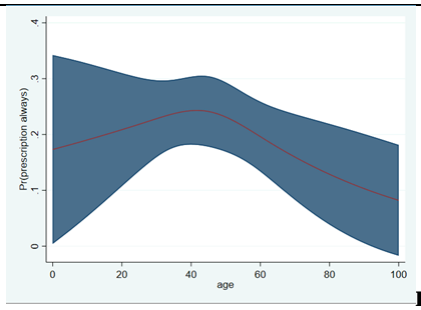
model 5



model 6



model 7



model 8

Table S6. Spline age graphs

The strength of IV is validated using F-statistics. We used the standard mean difference (SMD) to assess the covariate balance on the level of IV. The SMD based on the IV are generally less than 0.1 which can be treated as balanced covariate indicating IV is less likely to be associated with the unmeasured confounders.

Data	PIS	CPRD
F-statistics	857.01	449.973
SMD	Based on IV	Based on IV
Age	0.0073	0.0451
Gender: female	-0.0128	0.0059
Charlson comorbidity index _0	-0.0055	0.0836
Charlson comorbidity index _1	0.0081	-0.0289
Charlson comorbidity index _2	0.0012	-0.0506
Charlson comorbidity index _3	-0.0100	-0.0648
Charlson comorbidity index _4	-0.0014	-0.0396
socio-economic deprivation. 1	0.0512	0.0019
socio-economic deprivation. -2	-0.0059	0.003
socio-economic deprivation. -3	-0.0350	-0.0775
socio-economic deprivation. -4	-0.0025	-0.0328
socio-economic deprivation. -5	-0.0264	0.0884
Whether the monotonicity inequality hold	TRUE	TRUE

Table S7. Validation of IV assumptions. IVs are dichotomised at mean value.

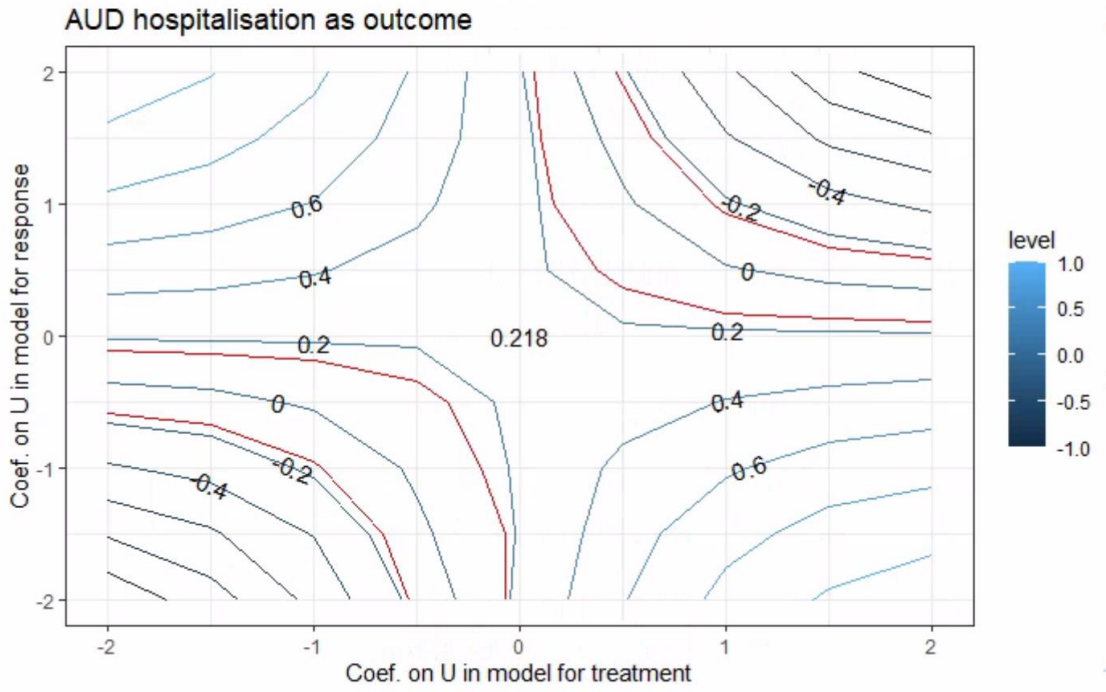


Figure S9. Sensitivity analysis to the unmeasured confounding. Results from the R function: survSensivity

Read Code	Description	Frequency of incident cases (%)
Eu10211	[X]Alcohol addiction	0.63
Eu10800	[X]Alcohol withdrawal-induced seizure	0.79
Eu10712	[X]Chronic alcoholic brain syndrome	0.04
Eu10212	[X]Chronic alcoholism	0.43
Eu10411	[X]Delirium tremens, alcohol induced	0.34
Eu10213	[X]Dipsomania	0.00
Eu10200	[X]Mental and behavioural disorders due to use of alcohol: dependence syndrome	0.26
Eu10300	[X]Mental and behavioural disorders due to use of alcohol: withdrawal state	0.03
Eu10400	[X]Mental and behavioural disorders due to use of alcohol: withdrawal state with delirium	0.02
E230.00	Acute alcoholic intoxication in alcoholism	1.63
E230z00	Acute alcoholic intoxication in alcoholism NOS	0.07
E230300	Acute alcoholic intoxication in remission, in alcoholism	0.02
E230000	Acute alcoholic intoxication, unspecified, in alcoholism	0.04
8H35.00	admitted to alcohol detoxification centre	0.16
E23..00	Alcohol dependence syndrome	52.51
E23z.00	Alcohol dependence syndrome NOS	1.23
E230.11	Alcohol dependence with acute alcoholic intoxication	0.03
Z191.00	alcohol detoxification	10.79
8BA8.00	alcohol detoxification	0.05
E010.00	Alcohol withdrawal delirium	0.21
E013.00	Alcohol withdrawal hallucinosis	0.07
E01y000	Alcohol withdrawal syndrome	8.02
G555.00	Alcoholic cardiomyopathy	0.52
J612.00	Alcoholic cirrhosis of liver	4.25
F11x011	Alcoholic encephalopathy	0.15
F394100	Alcoholic myopathy	0.06
F375.00	Alcoholic polyneuropathy	0.33
J671000	Alcohol-induced chronic pancreatitis	0.26
E23..11	Alcoholism	13.74
F11x000	Cerebral degeneration due to alcoholism	0.01
E012000	Chronic alcoholic brain syndrome	0.01
J617000	Chronic alcoholic hepatitis	0.05
E231.00	Chronic alcoholism	0.94
E231300	Chronic alcoholism in remission	0.02
E231z00	Chronic alcoholism NOS	0.65
E230100	Continuous acute alcoholic intoxication in alcoholism	0.00
E231100	Continuous chronic alcoholism	0.09
E010.12	Delirium tremens	0.43
E231.11	Dipsomania	0.02
E010.11	DTs – Delirium tremens	0.22
E230200	Episodic acute alcoholic intoxication in alcoholism	0.03
E231200	Episodic chronic alcoholism	0.11
E011000	Korsakov's alcoholic psychosis	0.28
E011100	Korsakov's alcoholic psychosis with peripheral neuritis	0.04
G852300	Oesophageal varices in alcoholic cirrhosis of the liver	0.18
E231000	Unspecified chronic alcoholism	0.10
C253.00	Wernicke's encephalopathy	0.18

Table S8. Read codes included for defining alcohol dependence.

Data request form for Chapter 5 (CPRD)

Administrator comments

2020-10-12 16:27: Study feedback: Amendments required

Comments:
Your application requires amendments to a section (or sections) before it may be approved. When amending these sections, please ensure you re-write the WHOLE section in the box, taking into account the reviewer/s comments. Please do not directly respond to reviewer/s comments in the box. Applications which do not re-write the whole section will be returned for further amendments.

2020-11-20 16:50: Amendments by Professor Jim Lewsey

General information

Study title Effectiveness and cost-effectiveness of alcohol use screening tests and treatments for alcohol-use disorders

Research area Drug Effectiveness, Economics, Pharmacoeconomics, Pharmacoepidemiology

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

Does this protocol involve requesting any additional information from GPs, or contact with patients? No

Research team

Applicant's role Role: Chief investigator and corresponding applicant
Email: jim.lewsey@glasgow.ac.uk
Name: Professor Jim Lewsey
Statistical experience: Yes
Experience of handling large datasets: Yes
Experience of practicing in UK primary care: No

Will the applicant be analysing the data? (Laaksonen et al.)

Error:This field is required for submission.

Collaborators

Collaborator's email: bhautesh.jani@glasgow.ac.uk
Will this person be analysing the data?: Yes
Status: Confirmed
Name: Dr Bhautesh Jani
Statistical experience: Yes
Experience of handling large datasets: Yes
Experience of practicing in UK primary care: Yes

Collaborator's email: francesco.manca@glasgow.ac.uk
Will this person be analysing the data?: Yes
Status: Confirmed
Name: Mr Francesco Manca
Statistical experience: Yes
Experience of handling large datasets: Yes
Experience of practicing in UK primary care: No

Collaborator's email: xxxxxxxx@student.gla.ac.uk
Will this person be analysing the data?: Yes
Status: Confirmed
Name: Miss Lisong Zhang
Statistical experience: Yes
Experience of handling large datasets: Yes
Experience of practicing in UK primary care: No

Collaborator's email: claudia.geue@glasgow.ac.uk
Will this person be analysing the data?: Yes
Status: Confirmed
Name: Dr Claudia Geue
Statistical experience: Yes
Experience of handling large datasets: Yes
Experience of practicing in UK primary care: No

Collaborator's email: linsay.gray@glasgow.ac.uk

Will this person be analysing the data?: Yes

Status: Confirmed

Name: Dr Linsay Gray

Statistical experience: Yes

Experience of handling large datasets: Yes

Experience of practicing in UK primary care: No

Collaborator's email: Elise.Whitley@glasgow.ac.uk

Will this person be analysing the data?: Yes

Status: Confirmed

Name: Dr Elise Whitley

Statistical experience: Yes

Experience of handling large datasets: Yes

Experience of practicing in UK primary care: No

Collaborator's email: Janet.Bouttell@glasgow.ac.uk

Will this person be analysing the data?: Yes

Status: Confirmed

Name: Mrs Janet Bouttell

Statistical experience: Yes

Experience of handling large datasets: Yes

Experience of practicing in UK primary care: No

Collaborator's email: Vittal.Katikireddi@glasgow.ac.uk

Will this person be analysing the data?: Yes

Status: Confirmed

Name: Professor Srinivasa Vittal Katikireddi

Statistical experience: Yes

Experience of handling large datasets: Yes

Experience of practicing in UK primary care: No

Collaborator's email: eileen.kaner@newcastle.ac.uk

Will this person be analysing the data?: No

Status: Confirmed

Name: Professor Eileen Kaner

Statistical experience: No

Experience of handling large datasets: No

Experience of practicing in UK primary care: No

Collaborator's email: Frederick.Ho@glasgow.ac.uk

Will this person be analysing the data?: Yes

Status: Confirmed

Name: Dr Frederick Ho

Statistical experience: Yes

Experience of handling large datasets: Yes

Experience of practicing in UK primary care: No

Access to data

Sponsor

Sponsor: University of Glasgow

(Sponsor information is retrieved automatically as the chief investigator's affiliation)

Funding source for the study

Is the funding source for the study the same as Chief Investigator's affiliation? Yes

Institution conducting the research

Is the institution conducting the research the same as Chief Investigator's affiliation? Yes

Method to access the data

Indicate the method that will be used to access the data Institutional multi-study licence

Is the institution the same as Chief Investigator's affiliation? Yes

Extraction by CPRD

Will the dataset be extracted by CPRD No

Data processors

Data processor is: Same as the chief investigator's affiliation

Processing: Yes

Accessing: Yes

Storing: Yes

Processing area: UK

Information on data

Primary care data CPRD GOLD, CPRD Aurum

Do you require data linkages Yes - I do require data linkages

Patient level data HES Admitted Patient Care, Mental Health Services Data Set (MHSDS), ONS Death Registration Data

Area level data

Do you require area level data? Yes

Practice level (UK) Practice Level Index of Multiple Deprivation

Patient level (England only) Patient Level Index of Multiple Deprivation

Withheld concepts

Are withheld concepts required? No

Linkage to a dataset not listed

Are you requesting a linkage to a dataset not listed? No

Patient data privacy

Does any person named in this application already have access to any of these data in a patient identifiable form, or associated with an identifiable patient index? No

Protocol information

Lay summary In general practices, there are various tests GPs use to assess whether their patients are at high risk of health harms because of the amount of alcohol they drink. If an individual's health starts to deteriorate because of their alcohol use the two major approaches to treatment available to GPs (or other health professionals) are to prescribe medications and/or to provide brief advice. However, evidence on what works is conflicting. The overall aim of this project is to examine how well do tests and treatments for risky drinking of alcohol work in a general practice setting and whether they provide value for money.

Analysing large datasets like CPRD provides an opportunity for finding out what interventions and treatments work and has some specific benefits, such as finding out how well things work in real-world circumstances. In this research we will study the effects of different tests and treatments for risky drinking by taking advantage of GPs tendencies to use the same testing tools or to prescribe the same treatment to their patients who have similar problems. For example, when there are a few options available for medicines to help reduce risky drinking, one GP may tend to prescribe a particular medicine more often than another.

We will calculate how much it costs to provide these tests and treatments and compare these costs against the likely health benefits over the short- and long-term. This will allow us to see which tests and treatments provide the best value for money.

Technical summary In this cohort study we will ascertain two cohorts of individuals and follow them up to quantify risk of future outcomes (primary care visits, hospitalisations and deaths) and to estimate remaining (quality adjusted) life expectancy.

Cohorts:
Cohort 1 (secondary prevention population): individuals at risk of alcohol harm (defined by screening tests such as AUDIT and FAST; self-reported ‘high’ level of alcohol consumption; other alcohol consumption related Read/Snomed codes)

Cohort 2 (tertiary prevention population): individuals hospitalised for any alcohol related condition (defined by ICD codes)

Screening test / treatment variables:
Alcohol use screening tests; Alcohol brief interventions; Pharmacological interventions (e.g. Acamprosate, Disulfiram, Nalmefene, Naltroxone)

Outcome variables:
Primary care record of alcohol use disorder; hospitalisations (alcohol intoxication / harmful use; alcohol dependency; alcoholic liver disease; liver disease (all)) and deaths (same categories as hospitalisations and all cause deaths)

Other variables (not mentioned above):
Patient's age, patient's sex, socio-economic deprivation (area-based), lab test results, comorbidity measured by Read/Snomed/ICD codes (e.g Charlson index)

Statistical methods:
Comparative effectiveness analyses for head-to-head comparisons of screening tests, alcohol brief interventions and pharmacological interventions will be carried out using multivariable logistic/Cox regressions, propensity score adjusted logistic/Cox regressions and instrumental variables adjusted logistic/Cox regressions. The latter will use physician's prescribing preferences and general practice preferences as instruments.

Decision analytic model / economic evaluation tool:
The outcomes from the statistical analyses will be used to populate a decision analytic model which will extrapolate the outcomes for the cohorts ascertained above using parametric survival modelling, validating against external data sources (e.g. national life tables). The different cohorts will allow economic evaluation of 'secondary' and 'tertiary' prevention strategies.

Sensitivity analyses:
As uncertainty exists in all aspects above, pre-specified sensitivity analyses will be undertaken.

Outcomes to be measured • Primary care record of alcohol use disorder (a. identified by AUDIT PC > 4; b. identified by AUDIT C > 4; c. identified by AUDIT C > 10; d. identified by FAST > 2; e. identified by 'Single question alcohol use test' (M-SASQ) > 1)
• Hospitalisation for alcohol intoxication / harmful use (AIH)
• Hospitalisation for alcohol dependency (AD)
• Hospitalisation for alcoholic liver disease
• Hospitalisation for liver disease (all)
• Death (caused by AIH)
• Death (caused by AD)
• Death (caused by alcoholic liver disease)
• Death (caused by liver disease – all)
• Death (all cause)

Note: the rationale for using alcohol use disorders (intoxication / harmful use / dependency) and alcoholic liver disease is that they make up a quarter of alcohol-attributable mortality, and are 100% alcohol attributable [1]

For developing the decision analytic model, we require linked data to all hospitalisation records.

Objectives, specific aims and rationale The overall objective of this research is to assess the effectiveness and cost-effectiveness of screening tests for alcohol harm and treatments for alcohol use disorders. To achieve this, we will carry out these specific aims:

1. calculate the comparative effectiveness of screening tests in primary care for alcohol harm and treatments for alcohol use disorders / alcohol-related outcomes using instrumental variables, propensity score matching and traditional observational approaches and compare the results from each method.

2. estimate the cost-effectiveness of screening tests in primary care and treatments for alcohol-use disorders using a decision analytic model developed using CPRD data sources.

Study background Drinking excessively is a major risk factor that affects health in the UK.

• NICE recommend alcohol screening to prevent alcohol-related outcomes. Five such screening tests are listed in current Public Health Guidance [2]. Little is known about the comparative effectiveness of these tests (e.g. AUDIT vs. FAST).

• Recent systematic reviews on pharmacologically controlled drinking [3] and pharmacotherapy for AUDs in outpatient settings [4] that conclude no high-grade evidence exists and reveal a lack of head-to-head comparisons in the literature.

• Alcohol brief interventions (ABIs) are structured conversations about alcohol consumption carried out between GP and patients. There is research suggesting that ABIs play an important role in reducing alcohol consumption among people who drink in harmful way, but not for alcohol dependent people, and is cost-effective [5]. However, recent trial evidence does not support effectiveness [6]

The above shows there is an evidence gap for (comparative) effectiveness of screening tests for alcohol harm and treatments (pharmacological and ABIs) for alcohol use disorders. Using CPRD data sources for our planned study provides a large 'real world population'. This addresses the well-known limitation of representativeness of clinical trials, and the large sample sizes are required to conduct high-quality instrumental variable analyses for addressing unmeasured confounding / confounding by indication.

Study type Hypothesis testing study

Study design Cohort study

Feasibility counts Epidemiology of alcohol dependence using CPRD has been previously carried out and published [7]. Comparative effectiveness and economic evaluation is feasible using this data source. As outlined in section J, we expect to be using health records from millions of patients (approximately 6.4 and 17.0 million from GOLD and Aurum, respectively). This will ensure counts of outcome events will be sufficiently large to support the multivariable regression models we will run (i.e. very confident that models with many covariates, 'using up' many degrees of freedom, will reach convergence when fitted).

Sample size considerations The large data set will support multivariable regression models with large number of covariates for all cohorts and for all outcomes described, as well as the instrumental variable analysis. The power for each pairwise treatment comparison will vary by how prevalent screening tests and treatments are, type of outcome, etc. However, given the large sample sizes in this study (exposure based on Read codes from [8] is approximately 6.4 and 17.0 million patients in GOLD and Aurum, respectively) we are confident that power will be very high for all pairwise comparisons. To further illustrate this, in a recent systematic review of randomised trials that compared the effectiveness of ABIs with no intervention for a quantity of alcohol drinking at 12 months outcome, the combined sample size from 34 trials was 15,197 [5].

Planned use of linked data and/or withheld concepts We plan to link CPRD to prescribing records, inpatient hospitalisations, mental health records and deaths. This is required to:

a) estimate comorbidities with as much relevant information as possible. The richer the information to construct models covariates, the better quality our comparative effectiveness research will be.

b) fully capture relevant health economic outcomes and resource use costs that are related to alcohol consumption to create a high-quality decision analytic model.

For our comparative effectiveness and decision analytic modelling, after we have identified the CPRD individuals in cohorts 1 and 2 we request all their HES inpatient records (regardless of ICD codes), primary care records, and mental health records (MHDS) before and after the index date when the individuals enter cohort 1 or cohort 2.

We feel this is justified because we will add high-quality evidence on effectiveness and cost-effectiveness of alcohol screening tests and treatments which will inform public health decision-making, clinical guidelines and ultimately the health of the UK population.

We are aware that as we are requesting a large amount of data, it will be be important to complete the data minimisation spreadsheet in due course so we only request the variables that we need. Here we provide an overview of the type of variables we require (and for what reason):

CPRD (GOLD and Aurum): patient variables (patient identifier for identifying linked records; age and sex will be covariates; registration deatils for cohort identification), practice variables (practice identifier and region will be needed to construct instrumental variables), consultation variables (dates, type of consultation needed for costing to inform decision analytic modelling), clinical variables (dates, type of event needed for costing to inform decision analytic modelling), referral variables (dates, type of referral needed for costing to inform decision analytic modelling), test variables (dates, category of event needed for costing to inform decision analytic modelling), all therapy variables (needed to identify treatment variables and also for costing to inform decision analytic modelling).

HES APC: patient variables (patient identifier for identifying linked records), all hospitalisation variables (exact dates needed for survival analysis modelling; method and source of admission needed for costing to inform decision analytic modelling), all episodes variables (exact dates needed for survival analysis modelling; method and source of admission needed for costing to inform decision analytic modelling), procedures variables (date of admission and date of discharge needed to calculate length of stay for costing to inform decision analytic modelling; OPCS codes needed for costing to inform decision

analytic modelling), all augmented care variables (for costing to inform decision analytic modelling), all critical care variables (for costing to inform decision analytic modelling), all health resource group variables (for costing to inform decision analytic modelling).

MHDS: we intend to use this data set only for calculating resource use and associated costs to inform decision analytic modelling. The types of variables we require are accommodation details, diagnsoses, specialty codings, occupation codes, services codings, consultation medium, admission method, discharge method, source of referral, status of service request.

The comparative effectiveness testing of screening tools and treatments for alcohol use disorders will benefit patients in England and Wales as it will increase the evidence base for clinical decision making. At present, there is unclear evidence on what screening tools and what treatments work best in community settings and our research will inform that. The findings on the appropriateness of using instrumental variables, propensity score matching and traditional regression approaches in CPRD data will benefit patients indirectly or over the longer term as it will facilitate (if findings favourable) further research. The cost-effectiveness findings will inform clinicians and policy-makers about the relative cost-effectiveness of screening tools and treatments and may influence future guidelines or investment in treatments. Better use of finite resources will improve patient care as the most effective care will be provided to patients.

Definition of the study population Our first cohort are those at risk of alcohol harm and subsequent disease and are potential targets for secondary prevention [9], whereas the second cohort have been hospitalised for an alcohol related condition and are potential targets for tertiary prevention [9].

Cohorts:
Cohort 1: : individuals at risk of alcohol harm (identified by screening tests such as AUDIT and FAST; self-reported ‘high’ level of alcohol consumption; other alcohol consumption related Read/Snomed codes) between 01/01/2004 and 31/12/2020 (or last available extraction time point). The index date for entry into the cohort will be earliest date from: screening test / high level of alcohol consumption / other alcohol consumption related Read/Snomed codes. Patients who are and who are not eligible for linked data to be included into cohort. The earliest of CPRD death date, transfer out date, and the end of study date will be used in

defining the end of follow-up. Only up-to-standard follow-up to be considered.

Provisional code list for inclusion into Cohort 1 (based on Read codes): see list in Supplementary Table 1.

Cohort 2: individuals hospitalised for any alcohol related condition (identified by ICD codes for alcohol intoxication / harmful use, alcohol dependency, liver disease (all)) between 01/01/2004 and 31/12/2020 (or last available extraction time point). The earliest of ONS linked death date, transfer out date, and the end of study date will be used in defining the end of follow-up. First/index hospitalisations will be coded by using a ‘look-back’ / screening period of 10 years (same fixed duration of 10 years look-back applied for each hospitalization). Cohort 2 will be a subset of Cohort 1.

Selection of comparison group(s) or controls The comparison groups for alcohol use screening tests will be, in turn, the most common tests, such as AUDIT and FAST.

The comparison groups for pharmacological treatments will be, in turn, the most common prescriptions for AUD, such as Acamprosate and Disulfiram.

The comparison group for ABIs will be no ABI administered.

Exposures, outcomes and covariates Exposures, outcomes and covariates
Screening test / treatment variables (Exposures):
Alcohol use screening tests (identified by Read codes [8]); pharmacological interventions (e.g. Acamprosate, Disulfiram, Nalmefene, Naltroxone); alcohol consumption (identified by Read codes [8], alcohol brief interventions (identified by Read/Snomed codes [8]). Notes: the comparisons in screening test, alcohol consumption and alcohol brief intervention exposures will be between different levels of these variables (including 'none'); the comparisons in pharmacological interventions will be head-to-head comparisons (e.g. Acamprosate vs. Disulfiram); screening tools will not be exposures for Cohort 2.

Note: the Read codes from [8] are available in ‘Supplementary Table 1’ and available for download from https://doi.org/10.17037/data.00001071. We have reproduced them in the Appendix of this application.

Instrumental variables: these will be determined by measuring the percentage of time specific GPs / GP practices (depending on level of aggregation of analysis) have prescribed / administered screening tools / treatments for patients with similar characteristics to the current one. The performance/validity of different instruments based on different lengths of 'history' (e.g. last 1/5/10/20 patients) will be tested.

Outcomes:
Primary care record of alcohol use disorder, hospitalisation for AIH, hospitalisation for AD, hospitalisation for alcoholic liver disease, hospitalisation for liver disease (all), death (caused by AIH), death (caused by AD), death (caused by alcoholic liver disease), death (caused by liver disease – all), death (all cause).

Covariates:
Age, sex, socio-economic deprivation, comorbidities (e.g. Charlson score)

Data/statistical analysis Statistical analysis plan:
1. Identify cohort 1 and cohort 2 by data science-type skills/coding (merging/ reshaping data, creating new variables needed for analysis)
2. Identify most common tests/treatments in data
3. Identify the two-way comparisons for statistical inference
4. Run comparative effectiveness analyses (a: unadjusted, b: adjusted (multivariable regression), c: adjusted (time-varying propensity score [10]), d: instrumental variable (e.g. physicians' prescribing preferences [11])
5. Develop decision analytic model for conducing economic evaluation of (causal) treatment effects identified in 4. We will undertake probabilistic sensitivity analysis (PSA) as part of this modelling. PSA involves specifying a distribution for each parameter included in the model. In order to characterise the uncertainty around the estimates of cost-effectiveness, the model will be run repeatedly to produce a range of estimates of cost and effectiveness. Each time the model is run, a different random value for each parameter will be selected from the distribution. This means that the uncertainty in every parameter is simultaneously taken into account.

We will develop an a priori modelling plan protocol before getting sight of the CPRD data. This will be subject to change after we become familiar with the exact nature of the CPRD

data set (we will version control this protocol).

Expanding on step 4 above, we will use Cox regression to model time to event of our outcome variables. Multivariable regression and time-varying propensity scores will be used to adjust for confounding variables that are measured in the data set (e.g. age, sex, socio-economic deprivation, comorbidities). A candidate propensity score technique will be inverse probability weighting based on pre-exposure characteristics (age, sex, socio-economic deprivation, comorbidities), building a score reflecting the probability of an individual being in the treatment group. This probability is then used in every regression as a weight for every individual. To address unmeasured confounding, we will implement multivariable instrumental variable Cox regression model, incorporating two-stage regression. First stage: using logistic regression of treatment on instrument to estimate the probability of receiving treatment A including all potential covariates (age, sex, socio-economic deprivation, comorbidities). Second stage: use Cox regression to estimate the risk of outcome using the probability from the first stage as exposure with the same list of covariates. The latent variable physician's prescribing preference (an example of an instrumental variable for the pharmacological interventions exposure) will be measured using the proxy of the most recent prescription made by the same physician. In terms of sensitivity analysis, we will run the same model using prior 2 to prior 20 prescriptions made by the same physician as instrumental variables. The results from instrumental variable models will be triangulated with the multivariable Cox regression and propensity score results.

The data sets for this project will be very large (many millions of patient records). We will develop our statistical code using a random subset of patients before submitting final code utilising high performance computer clusters at the University of Glasgow.

We will use both GOLD and Aurum in the above statistical analysis plan. Where we identify duplicate practices we will remove them from the GOLD data set. We will build our models using GOLD and use Aurum for validation purposes.

Plan for addressing confounding In this study, we use instrumental variable methods to tackle bias caused by unmeasured confounding. We will use the Durbin-Wu-Hausman approach to test the validity (H0: no endogeneity) of our constructed instruments by comparing ordinary least squares and instrument variable estimates.

Plans for addressing missing data We plan to use multiple imputation with chained equations procedures to address missing data under a missing at random (MAR) assumption. We will assume MAR rather than missing completely at random (MCAR) because in our experience of analysing similar routine healthcare data sets to CPRD (e.g. PIS and SMR in Scotland), MCAR is extremely unlikely as missing data is usually related to observed covariates. Although we cannot test missing not at random (MNAR) empirically, we will also consider the plausible mechanisms of any missing data in key variables.

Patient or user group involvement We have identified and are engaging with a patient group from another alcohol epidemiology/comparative effectiveness project - TRends and Inequalities in Prescribing for Alcohol use Disorders in Scotland (TRIPADS) – funded by Alcohol Change UK. We will explore the possibility of using this group for PPI in this project.

Plans for disseminating and communicating study results The study results will be published in peer-reviewed journals. No restriction on the extent and timing of publication.

Conflict of interest statement None.

Limitations of the study design, data sources, and analytic methods The main limitation is that there is a possibility that we will not identify valid instruments and if this happens our comparative effectiveness results could be biased due to unmeasured confounding.

References [1] Rehm J, Shield K. Global Burden of Alcohol Use Disorders and Alcohol Liver Disease. Biomedicines 2019; doi:10.3390/biomedicines7040099
[2] Public Health England. Guidance on the 5 alcohol use screening tests (1st June 2017). https://www.gov.uk/government/publications/alcohol-use-screening-tests/guidance-on-the-5-alcohol-use-screening-tests
[3] Palpacuer C et al. Pharmacologically controlled drinking in the treatment of alcohol dependence or alcohol use disorders: a systematic review with direct and network meta-analyses on nalmefene, naltrexone, acamprosate, baclofen and topiramate. Addiction 2017; doi: 10.1111/add.13974.

[4] Jonas D et al. Pharmacotherapy for Adults With Alcohol-Use Disorders in Outpatient Settings: a systematic review and meta-analysis. JAMA 2014; doi: 10.1001/jama.2014.3628.
[5] Kaner, E. F., Beyer, F. R., Muirhead, C., Campbell, F., Pienaar, E. D., Bertholet, N., ... & Burnand, B. (2018). Effectiveness of brief alcohol interventions in primary care populations. Cochrane database of systematic reviews, (2).
[6] McCambridge, J, and Saitz R. Rethinking brief interventions for alcohol in general practice. BMJ 356 (2017): j116
[7] Thompson A, Wright AK, Ashcroft DM, van Staa TP, Pirmohamed M. Epidemiology of alcohol dependence in UK primary care: Results from a large observational study using the Clinical Practice Research Datalink. PLoS One. 2017;12(3):e0174818. doi:10.1371/journal.pone.0174818
[8] Mansfield K, et al. Completeness and validity of alcohol recording in general practice within the UK: a cross-sectional study. BMJ Open 2019;9:e031537. doi:10.1136/bmjopen-2019-031537
[9] Winett, R. A. (1995). A framework for health promotion and disease prevention programs. American Psychologist, 50(5), 341–350. https://doi.org/10.1037/0003-066X.50.5.341
[10] Hernán, M. A., & Robins, J. M. (2006). Instruments for causal inference: an epidemiologist's dream? Epidemiology, 360-372.
[11] Davies, N. M., Gunnell, D., Thomas, K. H., Metcalfe, C., Windmeijer, F., & Martin, R. M. (2013). Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. Journal of clinical epidemiology, 66(12), 1386-1396

List of appendices

isac-protocol-application-form---e-and-ce-of-screening-and-treatments-for-aud_appendix.pdf

Supplementary material for Chapter 6

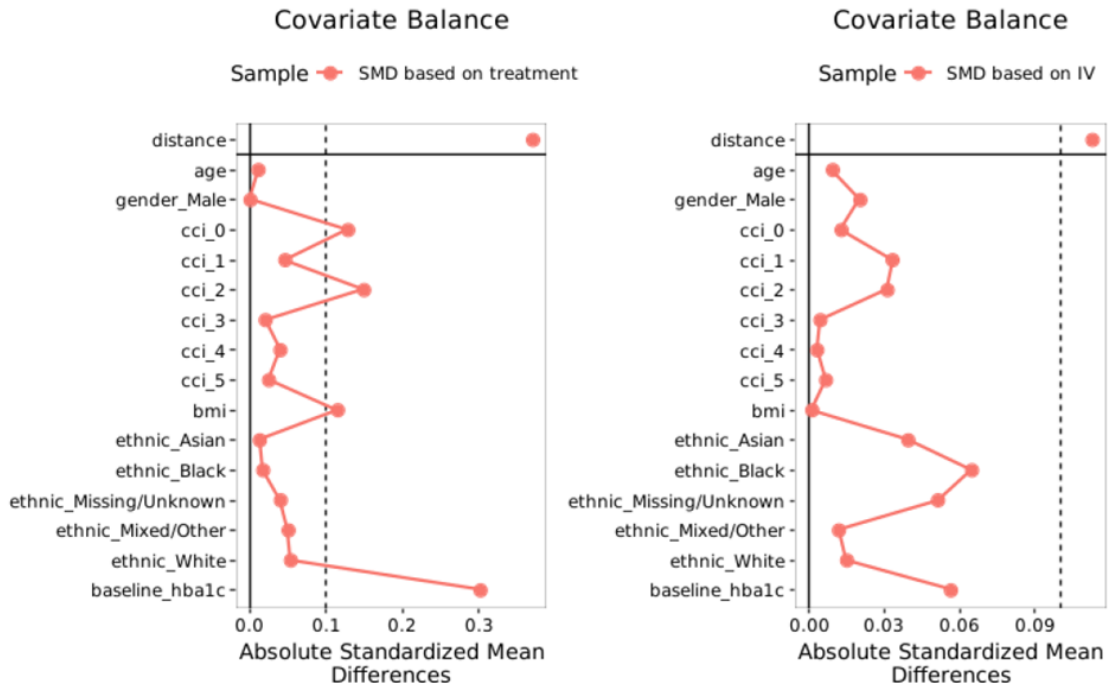


Figure S10. SMD of covariates based on the level of treatment (left), SMD of covariates based on the level of IV (right). The dash line represents represent SMD equals 0.1.

Weak instrument test (IV: the proportion of SU prescribed during last 1 year prescriptions, dichotomised at median)				
F-statistics	101.765		p value: <0.05	
Wu-Hausman test	1.581		p value: 0.209	
Validation of monotonicity assumption				
	Outcome: HbA1c reduced to between 42 mmol/mol and 48 mmol/mol		Outcome: HbA1cc reduced to less than 42 mmol/mol	
IV	IV inequality hold	Monotonicity inequality hold	IV inequality hold	Monotonicity inequality hold
The proportion of SU prescribed during the last 200 days prescriptions (dichotomised at median)	TRUE	TRUE	TRUE	TRUE
The proportion of SU prescribed during the last 1 year prescriptions (dichotomised at median)	TRUE	TRUE	TRUE	FALSE

Table S9. Validation of monotonicity assumption

```
Sensitivity Analysis to Unobserved Confounding

Model Formula: follow_hba1c ~ txA + age + gender + ethnic + cci + bmi + hba1c.x

Null hypothesis: q = 1 and reduce = TRUE
-- This means we are considering biases that reduce the absolute value of the current estimate.
-- The null hypothesis deemed problematic is H0:tau = 0

Unadjusted Estimates of 'txA':
  Coef. estimate: 2.5551
  Standard Error: 0.7752
  t-value (H0:tau = 0): 3.2959

Sensitivity Statistics:
  Partial R2 of treatment with outcome: 0.0029
  Robustness Value, q = 1: 0.0522
  Robustness Value, q = 1, alpha = 0.05: 0.0215
```

Figure S11. Sensitivity analysis to unmeasured confounding. Results from the R function 'sensemakr'.

Data request form for Chapter 6

Request for access to Scottish Diabetes Registry to predict long-term events and budget impact of alternative choices of drug treatments for patients with type 2 Diabetes Mellitus (T2DM)

Richard Grieve (LSHTM), Jim Lewsey and David McAllister (University of Glasgow)

Background

The proposed research will build from the NIHR-funded PERMIT study (see PERMIT | LSHTM) which aims to assess the long-term effectiveness and budget impact of alternative second-line drug treatments for patients with Type 2 Diabetes Mellitus. The study applies an instrumental variable (IV) design to Clinical Practice Research Datalink (CPRD) data to estimate the relative effectiveness and cost of alternative second-line treatments. These estimates are for the purpose of populating a microsimulation model (RAPIDS), to predict the rate of long-term complications, and accompanying budget impact.

The study's objectives are:

1. To assess short-term relative effectiveness according to individual risk factor profiles.
2. To calibrate and extend a microsimulation model developed in the US, for patients with T2DM in the UK (RAPIDS-UK).
3. To use the findings from objective 1, together with the model developed in objective 2, to estimate long-term effectiveness according to individual risk factor profiles, and project the NHS budget impact of personalising drug choice.

Rationale for request to Scottish Diabetes Registry

To fully address *Objective 1* requires that the instrumental variable approach is replicated on an external dataset. *Objective 2* of the PERMIT requires that the RAPIDS model is calibrated to relevant UK populations. Our request is therefore to calibrate the RAPIDS model using information from the Scottish diabetes register. We would then use this calibrated model to predict the budget impact to Scotland of alternative second line drug treatments for patients with T2DM (*Objective 3*).

The research would be undertaken by Lisong Zhang, a PhD student at the University of Glasgow, supervised by Jim Lewsey and David McAllister and if desired a member of the Edinburgh team. Lisong's PhD research is on using physician's prescribing preferences as an IV to adjust for unmeasured confounding in comparative effectiveness studies using observational data.

In extending objective 1 of the PERMIT study, Lisong will explore deriving instruments at different levels of aggregation (e.g. prescriber versus primary-care practice). In extending objective 2, Lisong would use information from the Scottish Diabetes Registry on for example, baseline characteristics (e.g. age, HbA1C prior to second-line treatment), to predict long-term events (e.g hospitalisations related to micro- and macro-vascular complications). Once calibrated, the model would be used to predict the costs of long-term events as well as medication costs. The resulting predictions of budget impact will be compared to those from a state transition cohort model that was developed using SCI-DC data.

Anticipated outputs

We anticipate that the proposed research would lead to two chapters of Lisong's PhD thesis and to subsequent publications. The usual approach of the Scottish Diabetes Research Network will be used to identify potential collaborators/co-authors.

About the PhD student

Lisong Zhang is a PhD student at the University of Glasgow supervised by Jim Lewsey and David McAllister. Lisong's PhD research is on using physician's prescribing preferences as instrumental variables to adjust for unmeasured confounding in comparative effectiveness studies using observational data. As well as replicating the instrumental variable analyses undertaken in CPRD, for the purposes of her PhD Lisong will explore deriving instruments at different levels of aggregation (e.g. at GP practice level) and using different time windows for calculating instruments (e.g. prior two years and prior 6 months as well as 1 year window that is used in our research).

OUTLINE PROPOSAL FORM: Research in collaboration with or as part of the Scottish Diabetes Research Network epidemiology group

1. Applicants: Principal applicant:

Name: **Lisong Zhang**

Affiliation: **University of Glasgow**

Email: **xxxxxxxx@student. gla.ac.uk**

Telephone: **xxxxxxxxxxx**

Address: **1 Lilybank Gardens Glasgow G12 8RZ**

Co-applicants and institutions:

David McAllister, Jim Lewsey (University of Glasgow)

Richard Grieve, Patrick Bidulka, David Lugo-Palacios (London School of Hygiene and Tropical Medicine)

2.Project title *(no more than 120 characters with spaces)*

Predict long-term events and budget impact of alternative choices of drug treatments for patients with type 2 Diabetes Mellitus (T2DM)

Start date: 06/2021

End date: 06/2022

3.Funding:

Has the project been or will it be peer reviewed? Yes ☐ No ☐

If so, by what organisation?

NIHR

Funding: *(If not specified above)*

Has funding been sought? * Yes ☐ No ☐

What is the deadline for application to the funder?

***Please note that applications for funding must be reviewed PRIOR to submission to a funding body and should be received AT LEAST two weeks before the deadline for submission*.**

4. Data sources:

Please check the data sources that are requested for this proposal & give full details in your scientific outline:

Scottish Diabetes register data: ☐

Linkage to other data sources (specify) : ☐

5. Ethical approval:

Does the study have ethical approval from a recognised Institutional Review Board/Ethics Committee? Yes ☐ No ☐

If Yes, please append a copy of the approval.

If No, please specify arrangements for obtaining appropriate approvals:

Yes, this study which is led by London School of Hygiene and Tropical Medicine (LSHTM) and also uses Clinical Practice Research Datalink (CPRD) data has been approved by LSHTM Internal Ethics Committee (ID 21395) and the MHRA Independent Scientific Advisory Committee (ID 20_064).

6.

Scientific outline: Please provide a **1-2 page outline** of your proposal, highlighting the specific requirements of the project for Scottish Diabetes Register data specified above. Please include:

1) **Lay summary (up to 250 words)**

The proposed research will build from the NIHR-funded PERMIT study (see PERMIT | LSHTM) which aims to assess the long-term effectiveness and budget impact of alternative second-line drug treatments for patients with Type 2 Diabetes Mellitus. The study applies an instrumental variable (IV) design to Clinical Practice Research Datalink (CPRD) data to estimate the relative effectiveness and cost of alternative second-line treatments. These estimates are for the purpose of populating a microsimulation model (RAPIDS), to predict the rate of long-term complications, and accompanying budget impact. The main objectives of these research are:

4. assess short-term relative effectiveness according to individual risk factor profiles.
5. To calibrate and extend a microsimulation model developed in the US, for patients with T2DM in the UK (RAPIDS-UK).
6. To use the findings from objective 1, together with the model developed in objective 2, to estimate long-term effectiveness according to individual risk factor profiles, and project the NHS budget impact of personalising drug choice.

2) **List of investigators**

Lisong Zhang

3) **Short background/ introduction**

The study applies an instrumental variable (IV) design to estimate the relative effectiveness and cost of alternative second-line treatments.

4) **Hypothesis**

The instrumental variable (IV) design reduce the possible unmeasured confounding bias in this observational study.

5) **Data requested**
   Scottish Diabetes Registry

6) Details of data handling and procedures for linkage

   The data handling and data linkage will comply with Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) and Reporting of studies Conducted using Observational Routinely-collected Data (RECORD) guidelines.

7) **Analysis Plan**

To fully address Objective 1 requires that the instrumental variable approach is replicated on an external dataset. Objective 2 of the PERMIT requires that the RAPIDS model is calibrated to relevant UK populations. Our request is therefore to calibrate the RAPIDS model using information from the Scottish diabetes register. We would then use this calibrated model to predict the budget impact to Scotland of alternative second line drug treatments for patients with T2DM (Objective 3).

The research would be undertaken by Lisong Zhang, a PhD student at the University of Glasgow, supervised by Jim Lewsey and David McAllister and if desired a member of the Edinburgh team. Lisong's PhD research is on using physician's prescribing preferences as an IV to adjust for unmeasured confounding in comparative effectiveness studies using observational data.

In extending objective 1 of the PERMIT study, Lisong will explore deriving instruments at different levels of aggregation (e.g. prescriber versus primary-care practice). In extending

objective 2, Lisong would use information from the Scottish Diabetes Registry on for example, baseline characteristics (e.g. age, HbA1C prior to second-line treatment), to predict long-term events (e.g hospitalisations related to micro- and macro-vascular complications). Once calibrated, the model would be used to predict the costs of long-term events as well as medication costs. The resulting predictions of budget impact will be compared to those from a state transition cohort model that was developed using SCI-DC data.

8) Details of authorship and collaboration agreement

7.

Agreement:

Signature: Lisong Zhang

GitHub link for the R code used in this thesis

https://github.com/zhanglisong810-ls/PhD-thesis-R-code/blob/1b0ba45571cadb53308234ad00bbfa0152706288/R_code