A causal inference framework for comparative effectiveness and safety research using observational data, with application in type 2 diabetes

Doctoral Thesis

Laura Maria Güdemann

Submitted by Laura Maria Güdemann to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Medical Studies in August 2023

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Signature: ....

Abstract

Randomized controlled trials are the gold standard to answer causal questions in health research as the process of randomization ensures balanced treatment groups and therefore makes it possible to compare average group outcomes directly. But they have many limitations with respect to costs, ethical considerations and practicability and therefore may not be suitable to answer all research questions. Evidence on cause and effect relationships from observational studies have the potential to overcome the limitations of trials and close important research gaps as they provide the possibility to study subpopulations of patients which are often excluded due to safety concerns, or can give insights into the risk profile of long-term endpoints. The quality of this real-world evidence depends on the quality of data, their suitability to answer a particular research question and the use of appropriate methods to estimate the treatment effect of interest. Of concern in observational research is bias in the treatment effect estimation due to confounding, as the treatment assignment is not controlled by the researcher and cannot be randomized. It is therefore possible that treatment groups are not balanced and confounding factors exist in the data which influence the treatment choice and the outcome of interest simultaneously.

The benefits of observational studies make them attractive for studying the risk and benefit profiles of oral type 2 diabetes treatments, especially of newer agent classes such as Sodium-glucose Cotransporter-2 Inhibitors. Prescribing of this treatment class has increased in recent years and a large proportion of type 2 diabetes patients have become eligible to receive agents from this class after recent treatment guideline changes. More information about treatment effects of Sodium-glucose Cotransporter-2 Inhibitors are needed especially for the large patient population of older adults (e.g. 70 years or older), as possible adverse

effects such as osmotic symptoms associated with this class could have severe consequences for these patients.

The overall aim of this thesis is to develop a causal inference framework for the exploitation of observational data, needed to derive high quality evidence on the benefit and safety profile of oral type 2 diabetes treatments, with a focus on the widely prescribed treatment class of Sodium-glucose Cotransporter-2 Inhibitors and the patient population of older adults. Chapter 1 and 2 are introductions to causal inference theory including the description of all estimation methods employed in this thesis and an introduction to type 2 diabetes research encompassing important treatment decision considerations, and current research evidence on Sodium-glucose Cotransporter-2 Inhibitors. Chapter 3 presents a triangulation framework of assorted estimation methods to establish the consistency of estimation results from approaches utilizing different parts of the data and relying on different data structure assumptions. Furthermore, an Instrumental Variable approach is introduced which uses data from the period before treatment initiation to mitigate potential bias in case the exchangeability assumption is violated and a history of the outcome of interest previous to treatment initiation has an influence on the treatment decision. Chapter 4 describes a simulation study on the performance of established construction methods for a proxy Instrumental Variable of health care provider prescription preference. The methods are tested under different data conditions such as change in provider preference over time, missing data in baseline covariates and different sample sizes within each health care provider. Additionally, a construction method is introduced that aims to address changes in preference over time and non-ignorabile missingness in baseline characteristics. In Chapter 5 the developed conclusions about a robust Instrumental Variable estimation approach from previous chapters are applied for a causal analysis on the relative benefit and risk profile of Sodium-glucose Cotransporter-2 Inhibitors versus Dipeptidyl peptidase-4 Inhibitors in the patient population of older adults. Chapter 6 provides an overview of the main findings and implications of this thesis and discusses limitations and future research potential of each study.

Acknowledgements

I would like to express my deepest gratitude to my supervisors Dr Jack Bowden, Dr Beverley Shields, Dr John Dennis and Dr Lauren Rodgers for their exceptional support and guidance. It has been inspirational to work with you and I will always appreciate the time and effort you put into giving me constructive feedback and helping me develop throughout this academic journey.

Additionally, I would like to thank my colleagues and collaborators at the Exeter Centre of Excellence for Diabetes Research and the University of Exeter for their advice and help as well as their friendship. The welcoming and collaborative spirit of this team has been a comforting source of positivity, especially during the pandemic and during the final 'writing up' phase of this thesis.

I have been lucky to have made treasured friendships with fellow PhD students Jana Sönksen, Deniz Türkmen, Vasileios Karageorgiou and Mary Fredlund and I would like to thank them for all their emotional support and all the time they spent with me during our writing marathons and other adventures. I also would like to extend sincere thanks to my close and supportive circle of 'home' friends: Sonja Henn, Laura Würfel-Swetik, Anna von Keudell and Nele Käfer.

Nobody has been more important to me along this journey than my family. I am deeply thankful for my parents who have inspired my path in academia by filling my childhood with 'WAS IST WAS' books and my head with an 'I can do hard things' attitude. Their unwavering believe in me has been essential. I also would like to thank my wonderful siblings for always cheering me on along the way. Lastly, I want to express my heartfelt appreciation to Christopher and Pepper who have supported me everyday with their endless love, patience and encouragement.

Contents

List of Figures

List of Tables

List of abbreviations

AE	adverse effect
aT	as treated
ATE	average treatment effect
ALT	alanine aminotransferase
AME	Average marginal effect
BMI	body mass index
CaT	Corrected as treated
CF	Control Function
CI	confidence interval
CKD	chronic kidney disease
CPRD	Clinical Practice Research Datalink
CVD	cardiovascular disease
DAG	directed acyclic graph
DiD	difference-in-difference
DKA	diabetic ketoacidosis
GLP-1	glucagon-like peptide 1 agonists
DPP4i	Dipeptidyl peptidase-4 Inhibitors
EBM	evidence-based medicine
eGFR	glomerular filtration rate
EHR	electronic healthcare records
EMA	European Medicines Agency
FDA	US Food and Drug Administration
FN	False negatives
FNR	False negative rate
FP	False positives
FPR	False positive rate

GP	general practice
HbA1c	glycated haemoglobin
HR	hazard ratio
HES	Hospital Episode Statistics
IMD	Index of Multiple Deprivation
ITT	intention to treat analysis
IV	Instrumental Variable
IV alldichmean	IV based on all prescriptions (dichotomized with mean)
IV alldichmedian	IV based on all prescriptions (dichotomized with median)
IV allprop	IV based on all prescriptions (proportion)
IV ePP	IV constructed with the Ertefaie method
IV ePP (rirs)	IV constructed with our proposed extended Ertefaie method
IV(PP)	True simulated PP as IV, utilizing all data in case of missingness
IV(PP) cc	True simulated PP as IV, utilizing complete case data in case of missingness
IV prevpatient	IV based on previous prescription
IV prev2patient	IV based on previous 2 prescriptions
IV prev5patient	IV based on previous 5 prescriptions
IV prev10patient	IV based on previous 10 prescriptions
IV allprevprop	IV based on all previous prescriptions
IV star	IV constructed with the Abrahamowicz method
LATE	local average treatment effect
MAD	mean absolute difference
MCAR	missing completely at random
MFN	Metformin
MNAR	missing not at random
MSE	mean squared error
NPV	Negative predictive value
NUC	no unmeasured confounding
ONS	Office for National Statistics

Obs. estimate	Observational estimate, multivariable regression adjusted for measured confounders
PERR	prior event rate ratio
PP	prescription preference
PPV	Positive predictive value
POA-IV	prior outcome augmented Instrumental Variable approach
POA-CF	prior outcome augmented Control Function approach
PSM	propensity score matching
RCT	randomized controlled trial
RMSE	relative root mean squared error
RWE	real-world evidence
SE	standard error
SGLT2i	Sodium-glucose Cotransporter-2 Inhibitors
SU	Sulfonylureas
T1D	type 1 diabetes
T2D	type 2 diabetes
TN	True negatives
TNR	True negative rate
TP	True positives
TPR	True positive rate
TSLS	Two-Stage Least Squares
TZD	Thiazolidinediones

Notation

Variables:

Y	Health outcome of interest (binary or continuous)
X	Treatment decision (binary)
Z	Instrumental Variable
W	Measured confounder(s)
U	Unmeasured confounder(s)
R	Randomization
PP	Provider preference
$\hat{\Delta}$	Residuals for the Control Function approach
T	Time of prescription
V	Provider level influence on missing data
F	Missingness indicator
$\mathbf{O}$	Minimal variable set that satisfies the backdoor criterion if controlled for in the causal analysis

Population:

N	Study population size
N_{Tx}	Treatment group size
N_{Ct}	Control group size
J	Number of health care provider in the study population
$j = 1, \ldots, J$	Index of individual provider in the study population
n_j	Number of patients treated by provider j
$i = 1, \ldots, n_j$	Patients' index within provider j (ordered by T)

Treatment effects:

β	Treatment effect of interest
$\hat{\beta}_{aT}$	As Treated estimate
$\hat{\beta}_{CaT}$	Corrected as treated estimate

$\hat{\beta}_{\mathrm{PSM}}$	Treatment effect estimated with a propensity score matching model
$\hat{\beta}_{\mathrm{IV}}$	Treatment effect estimated with the Instrumental Variable method
$\hat{\beta}_{\mathrm{CF}}$	Treatment effect estimated with the Control Function method
$\hat{\beta}_{\mathrm{DiD}}$	Treatment effect estimated with the difference-in-difference method
$\hat{\beta}_{\mathrm{POA-IV}}$	Treatment effect estimated with prior outcome augmented Instrumental Variable method
$\hat{\beta}_{\mathrm{POA-CF}}$	Treatment effect estimated with prior outcome augmented Control Function method
$\widehat{\mathrm{HR}}_{\mathrm{PERR}}$	PERR estimate of the treatment effect (time-to-event data)
$\widehat{\mathrm{HR}}_{\mathrm{prior}}$	Hazard ratio of the prior outcome model
$\widehat{\mathrm{HR}}_{\mathrm{study}}$	Hazard ratio of the study outcome model

Other notation:

π	Propensity score
$^{\circ}$	indicates propensity score matched data
Q_{e}	Generalized heterogeneity statistic
b	Number of previous prescriptions used to construct the proxy instrument for provider preference
L_{j}	Number of period for the generation of continuous change of provider j
$\mathrm{i}^{\star}$	Change time of provider prescription preference

Chapter 1

Introduction to causal inference

1.1 Structure and aims of this thesis

The overall aim of this thesis is the development of a causal inference framework for the exploitation of observational data, needed to derive high quality evidence on the benefit and safety profile of oral type 2 diabetes treatments, with a focus on the widely prescribed treatment class of Sodium-glucose Cotransporter-2 Inhibitors and the large patient subpopulation of older adults, as these patients are under-represented in clinical trial studies.

Chapter 1 aims to introduce the theory of causal inference to estimate the causal effect of an exposure/ treatment of an health outcome of interest such as the achieved reduction of blood glucose levels or the experience of adverse effects. For this introduction, a formal definition of the causal effect using the counterfactual framework is given and directed acyclic graphs are established as a visualization tool utilized in this thesis to communicate data structure assumptions of causal inference methods made for the treatment effect estimation. Additionally, this chapter aims to reason the concept of integrating evidence of randomized controlled trials and observational studies for high quality evidence-based medicine. Similar to this concept is the triangulation of evidence from different estimation methods which is valuable to discuss the consistency of research evidence. Finally, this chapter aims to provide and overview of all estimation methods employed in this thesis' observational studies and their data structure assumptions.

Chapter 2 aims to provide an introduction to type 2 diabetes and considerations important for clinicians to make individualized treatment decisions. Furthermore, this chapter aims to illustrate limitations of randomized controlled trials and their implications to provide adequate evidence for treatment guidelines for type 2 diabetes, especially for the important subpopulation of older patients. A comprehensive summary of current evidence on the benefit and risk profile of Sodium-glucose Cotransporter-2 Inhibitors is provided. This discussion will highlight current research gaps on the use of Sodium-glucose Cotransporter-2 Inhibitors in older type 2 diabetes patients and ways to close them with causal evidence from observational data.

Chapter 3 aims to present the development of an approach for causal effect estimation which utilizes data from the periods before and after study treatment initiation and combines aspects of the difference-in-difference and Instrumental Variable approach. This approach aims to mitigate potential bias under data structure conditions which violate necessary assumptions of these two methods. Furthermore, this chapter aims to showcase a triangulation framework of assorted causal inference methods which leverage different parts of the data and rely on different data structure assumptions to judge the consistency of estimation results and identify potential sources of bias in observational data.

Chapter 4 aims to provide a comprehensive summary and state of the art estimation performance analysis of different Instrumental Variable methods using the proxy design for healthcare provider prescription preference as instrument. The different construction methods for a proxy variable of prescription preference are assessed under different data situations regarding sample size, missingness in confounder data and different structures of change in preference over time. Furthermore, this chapter aims to introduce an extended version of a construction and estimation method for a preference-based instrument which is able to accommodate change in prescription preference over time as well as non-ignorable missingness in measured confounders.

Chapter 5 aims to utilize the causal estimation framework established in previous chapters to investigate the benefit and risk profile of Sodium-glucose Co-transporter-2 Inhibitors in type 2 diabetes patients 70 years and older. This causal inference analysis aims to determine the relative benefit of this treatment class compared to Dipeptidyl peptidase-4 Inhibitors regarding the reduction of blood glucose levels and body weight as well as the relative risk of experiencing adverse effects which are particularly of concern in older patients such as genital infections, volume depletion/ dehydration or diabetic ketoacidosis.

Chapter 6 is a summary of the main results and discussion of the limitations and future research opportunities for each chapter.

Chapter 3, 4, 5 have been written as full manuscripts to be published in topical peer reviewed journals. Supplementary online material, such as analysis codes and data descriptions are available and will be directed to in the respective chapters.

1.2 Introduction

Medical research often tries to answer cause-and-effect questions, for example about factors causing a disease or influencing the risk of experiencing adverse effects. The gold standard for the analysis of causal questions are randomized controlled trials, which analyse the difference in an health outcome of interest between two groups of individuals exposed or unexposed to a study treatment. The process of randomizing the treatment assignment ensures that both treatment groups are comparable regarding individuals characteristics. Therefore, it is possible to compare the causal effect of the treatment on the outcome directly. [1, 2, 3] Randomized controlled trials are not practical or ethically acceptable to answer all medical research questions. [3, 4] In such cases, evidence from observational studies, for example analysing primary care patient records, insurance claims, or registers, can be valuable to overcome limitations of trials. [4, 5]

The value of these studies to answer causal questions lies in the quality of the data, their suitability to answer a particular research questions and the use of appropriate analysis methods. As treatment allocation in observational data is not randomized, confounding of factors influencing the treatment decision and the outcome of interest simultaneously is possible. If confounding is not accounted for, it can lead to bias in the estimation of the treatment effect and hence result in incorrect conclusions about the cause-and-effect relationship under investigation.

This chapter provides an introduction to causal inference theory, starting with the formal definition of the causal effect of interest, utilizing the counterfactual framework and directed acyclic graphs for the visualization of the assumed data structure under which the causal effect is studied. Directed acyclic graphs are a helpful tool which will be employed throughout this thesis to communicate data structure assumptions of treatment effect estimation methods. [4, 6, 7] In the following, a detailed explanation of confounding bias in observational research and an introduction to the idea of integrating evidence from randomized controlled trials and observational studies is given. Furthermore, the idea of triangulating evidence from estimation methods which use different parts of the data at hand, make different data structure assumptions and are potentially subject to different sources of bias is introduced. As the analysis of treatment effects in observational data is complex and may be influenced by confounding bias, triangulating results from different methods provides a possibility to assess the consistency of estimation results. [8, 9, 10, 11] Lastly, an overview of all methods for the estimation of treatment effects, employed in latter chapters of this thesis, is given. The methods differ in their ability to account for measured and unmeasured confounding. Special focus is placed on the Instrumental Variable method using a proxy variable for healthcare provider prescription preference as instrument in order to create a pseudo-randomized sample and minimize the risk of unmeasured confounding in the causal effect estimation. For each method important assumptions on the data structure are explained.

1.3 The concept of causality and causal inference

Defining questions in medical research and epidemiology concern cause-and-effect relationships such as:

- Which factors cause a disease;
- How does a certain therapy or treatment effect the duration/ course of the disease;
- What is the efficacy of a drug in a given population;

rather than questions of association. Inferring causality requires knowledge of the data-generating process and - different to associations - cannot be computed only from the data at hand. [12, 13, 14] Causal knowledge in medical research is valuable as it helps to understand what can be done or should be avoided in order to achieve desired outcomes (e.g. treatment targets) or avoid undesirable consequences (e.g. side effects). [15] Discussions about the precise definition of causality as a concept dates back to the 18th century. The philosopher David Hume offered a widely known definition of the concept stating that: 'We may define cause to be an object, followed by another, ... where, if the first object had not been, the second had never existed'. [16] This definition focuses on a specific aspect of causation in that event A causes event B, if the occurrence of event A was necessary for the occurrence of event B under the observed background circumstances. [6] A similar definition is given by Pearl [17] who states that 'A variable X is a cause of a variable Y if Y in any way relies on X for its value. . . X is a cause of Y if Y listens to X and decides its value in response to what it hears'. The process of causal inference uses data of a sample/ study cohort to infer these kinds of cause-and-effect relationships in the target population of interest. [13] It is argued that we are never able to prove causal relationships beyond any doubt from the impression of a sample or repeated samples as faulty observations can never be ruled out. Therefore, rather than proving causality, medical research relies on empirical evidence to strengthen arguments for a possible causal relationships between an exposure (e.g. treatment or medical intervention) and an outcome (e.g. disease). [18, 19]

1.4 Counterfactual framework and definition of the causal effect

Counterfactual or potential outcomes is a conceptual and notational framework and is useful for the analysis of causality and estimation of causal effects. It is commonly utilized in psychology, statistics and epidemiology and formalizes how humans naturally think about causality. [6, 4, 7] Throughout this thesis, the outcome of interest is denoted with Y. We are interested in defining the individual causal effect of a treatment X for individual i in a given population of size N. In its basic form, the framework assumes that individuals can be assigned to two alternative states of treatment, the index treatment $X = 1$ (e.g. intake of a specific drug) or the control $X = 0$ (e.g. placebo or no administration of drug). When comparing two active treatments in this framework they are sometimes referred to as treatment B and treatment A respectively. The treatment effect is defined for the index treatment compared to the control treatment. The assumption of a binary treatment will be maintained throughout this thesis. In this framework treatment assignment must be manipulable and each treated individual has a theoretically observable outcome for the active treatment ($Y_i^{X=1}$) and the control ($Y_i^{X=0}$). [4, 6] The definition of a causal effect measure can be of additive or multiplicative scale (if the outcome is strictly positive) and depends on the type of outcome variable of interest. [12, 20] For a continuous outcome, the causal effect of the treatment on the outcome for each individual, expressed on an additive scale, is:

$$\beta_i = Y_i^{X_i=1} - Y_i^{X_i=0}.$$

As only one of the treatment states for each i is observable, β_i is not directly measurable. Instead, the focus of causal inference is usually on estimation of the average causal effect for the population of interest:

$$\beta = E[Y|X = 1] - E[Y|X = 0]$$

This is the expected difference of potential outcomes in case every individual in the population receives the treatment and in case every individual receives the control. Individuals will only receive one of treatment alternatives and therefore causal inference methods focus on consistently estimating $E[Y|X = 1] = \bar{Y}^{X=1}$ and $E[Y|X = 0] = \bar{Y}^{X=0}$ from individuals in the treatment group ($i \in Tx$) and in the control group ($i \in Ct$). The standard estimator for the average treatment effect (ATE) in a study cohort is:

$$\hat{\beta} = \bar{Y}_{i \in Tx}^{X=1} - \bar{Y}_{i \in Ct}^{X=0},$$

where $\bar{Y}$ denotes the sample average. [4, 6] In order to estimate this effect measure with observable data, identifying assumptions about the data structure and underlying causal relationships are necessary. These assumptions are often not fully testable and therefore need to be justified with subject matter knowledge. Four identifying assumptions need to be considered:

1. **Exchangeability:** It requires individuals in both treatment groups to have on average the same potential outcome under assignment to treatment or control. That is: $\bar{Y}_{i \in Tx}^{X=1} = \bar{Y}_{i \in Ct}^{X=1}$ and $\bar{Y}_{i \in Tx}^{X=0} = \bar{Y}_{i \in Ct}^{X=0}$. This assumption is necessary in order to be able to utilize the observed outcomes of both treatment groups as proxy for their counterfactual unobserved potential outcomes. [21] **Conditional exchangeability** requires exchangeability to hold after conditioning on a set of measured variables.

2. **Positivity assumption:** It specifies that each individual must have a non-zero probability to receive either of the treatment choices. When conditioning on other variables, this assumption needs to hold for all possible co-variate combinations. A structural positivity violation is possible for example in case of contraindication of a treatment for certain patient characteristics. [20, 22]

3. **Consistency assumption:** This assumption is sometimes referred to as treatment variation irrelevance and states that the exposure of interest needs to be sufficiently well defined in order for each individual to have one potential outcome for each treatment condition. It is violated for example in case of multiple different versions of the treatment. [20, 23]

4. **Non-Interference assumption:** It requires that an individual's potential outcome is not dependent on the exposure status of other individuals in the population. Spillover effects of some exposures such as vaccinations for infectious diseases can lead to a violation of this assumption. [20, 24]

The consistency assumption and non-inference assumption together are often referred to as stable unit treatment value assumption. [4, 20]

If the exchangeability assumption does not hold and there exists systematic differences between the treatment groups, the standard estimator will not consistently estimate the true ATE in the population. There are two main reasons why exchangeability might not hold in the analysis. A restriction of the analysis to selected individuals can lead to selection bias if the selection process is affected by the treatment and outcome of individuals or their causes. A selection would be for example necessary if the loss of follow-up data was affected by side-effects of the study treatments or symptoms of the outcome of interest. Another source of bias due to inherent differences between study and control group can arise when the treatment and outcome share an uncontrolled common cause. This is referred to as confounding/ confounding bias and a confounder is a variable that affects the treatment assignment and the outcome simultaneously. Additionally, the distinction is made between measured and unmeasured confounders, referring to whether or not the variables are recorded in the data at hand. Often, the exchangeability assumption is also called 'no confounding' assumption. If the confounders are measured it is possible to control for them in the analysis. In this case the analysis will rely on the conditional exchangeability assumption or the assumption of 'no unmeasured confounding' (NUC). [6, 20, 25]

1.5 Causal diagrams and directed acyclic graphs

Causal effects can only be identified in the context of relationships of important components in the system related to the causal question at hand. Causal diagrams are a powerful graphical tool to represent prior knowledge and assumptions about causal mechanisms, such as statistical dependence, independence, and underlying causal effects of interest. [13, 20, 26] Figure 1.1 shows a causal diagram for the assumed relationships between a treatment X, outcome Y and a measured confounder W, for example a biomarker such as age or weight.

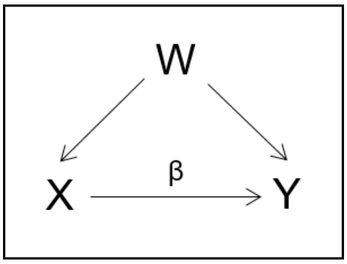
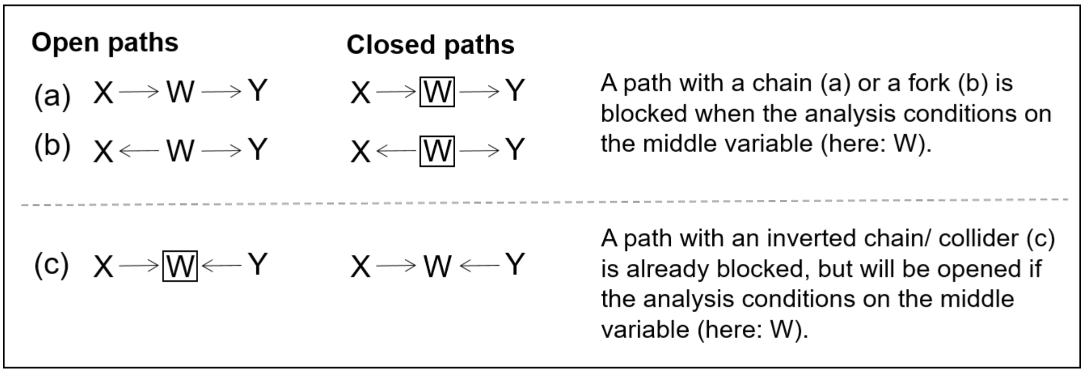


Figure 1.1: Example of a directed acyclic graph depicting the assumed relationship between the variable X*: treatment,* Y*: health outcome of interest and* W*: measured confounder, for example age or weight. We are interested in estimating the causal effect β of* X *on* Y*.*

Each variable in a causal diagram is represented at a specific point in time and referred to as a node. An arrow from variable X to Y indicates the assumption of a causal relationship between these two variables. In Figure 1.1 this is the average causal effect β of the treatment X on the health outcome Y we want to estimate. As each connection between nodes has an arrow of effect which is 'directed', causal diagrams which entail instantaneous non-cyclical relationships are also called directed acyclic graphs (DAGs). [27] Furthermore, a lack of an arrow between nodes indicates that no relationship between the variables is assumed. A path is a sequence of arrows that connects two variables. It can be causal or non-causal depending on the direction of the arrows. Specifically, if all arrows on a path point in the same direction, such as the path X → Y in Figure 1.1, each variable on the path causes the subsequent variable and the path is causal. Arrows on a non-causal path on the other hand do not all point in the same direction, such as the path X ← W → Y in Figure 1.1. Paths can also be blocked or unblocked depending on their structure and conditioning strategy of the analysis. Conditioning on a variable in the analysis is marked in a DAG with a rectangle around the respective variable. Three important graphical rules (known collectively as 'd-separation') determine whether a path is open or closed, a summary of these rules is given in Figure 1.2. [28]


Open paths
Closed paths
(a) $X \rightarrow W \rightarrow Y$
$X \rightarrow \boxed{W} \rightarrow Y$
(b) $X \leftarrow W \rightarrow Y$
$X \leftarrow \boxed{W} \rightarrow Y$
(c) $X \rightarrow \boxed{W} \leftarrow Y$
$X \rightarrow W \leftarrow Y$

A path with a chain (a) or a fork (b) is blocked when the analysis conditions on the middle variable (here: W).

A path with an inverted chain/ collider (c) is already blocked, but will be opened if the analysis conditions on the middle variable (here: W).

Figure 1.2: D-separation rules to decide which paths in a DAG are open and closed/blocked.

Paths a) and b) as depicted in Figure 1.2 represent a chain and a fork respectively. These paths are open and can be blocked if the analysis conditioned on the middle variable, here W. An example would be a regression model of Y on X with W as covariate. An inverted fork with a collider (W) is presented with path c). This path is blocked due to the collider, but will be opened if the analysis conditions on W.

Appropriately drawn DAGs and correct application of the d-separation rules helps to identify whether an observed association between X and Y has a causal interpretation or instead might be due to a spurious association. Additionally, it can be identified which variables the analysis should be conditioned on in order to consistently estimate the causal effect of interest. The "backdoor path criterion" is a tool to identify the set of variables $\mathbf{O}$ in a given DAG that need to be controlled for in the analysis in order to identify the causal relationship between X and Y. For this criterion to be satisfied no node in $\mathbf{O}$ can be affected by X and conditioning on all variables in $\mathbf{O}$ must ensure that all backdoor paths between X and Y are blocked. [27, 28] In Figure 1.1, X and Y are connected via the direct path X → Y and the backdoor path X ← W → Y. This backdoor path includes a fork and can be blocked by conditioning on the confounder W in the analysis, if this variable has been measured in the data. If β is estimated without controlling for W the observed association between X and Y can not be interpreted as causal and might be due to spurious association, confounding bias. Figure 1.3 represents a more complex DAG including 5 measured confounders (W_g and $g = 1, \ldots 5$) and an unmeasured confounder U. In order to determine if it is possible to estimate the causal effect of X on Y with the data at hand and the assumed relationships, the backdoor path criterion can be applied. Besides the direct path X → Y there are four backdoor paths from X to Y and the d-separation rules (Figure 1.2) help decide which of them need to be blocked in order to estimate the causal effect without bias.

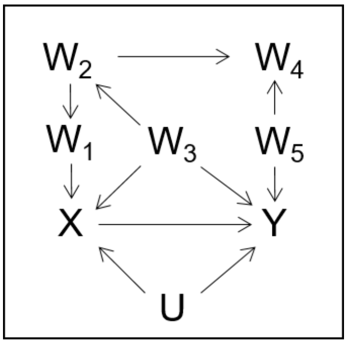


Figure 1.3: DAG representing a complex data structure between X: treatment, Y: outcome of interest, $W_1, \ldots, W_5$*: measured confounders and* U*: unmeasured confounder.*

The backdoor path $X \leftarrow W_3 \rightarrow Y$ includes a fork and can be blocked if the analysis conditions on the confounder W_3. This will also block the backdoor paths $X \leftarrow W_1 \leftarrow W_2 \leftarrow W_3 \rightarrow Y$ and $X \leftarrow W_3 \rightarrow W_2 \rightarrow W_4 \leftarrow W_5 \rightarrow Y$, which are defined by chains and a fork and would be otherwise open. The fourth backdoor path $X \leftarrow W_1 \leftarrow W_2 \rightarrow W_4 \leftarrow W_5 \rightarrow Y$ is already blocked as it contains a collider W_4. Conditioning on W_4 will open the path and introduce collider bias in the causal effect estimation. Lastly, the backdoor path X ← U → Y is open and cannot be blocked, as it is not possible to condition on a variable that is not measured in the data. In summary, the backdoor path criterion cannot be fulfilled with the data at hand and relative to the assumed data structure, because $\mathbf{O} = (\mathsf{W}_3, \mathsf{U})$ entails an unmeasured confounder. As result, confounding bias due to the spurious association introduced in the causal effect estimation by the open backdoor path including U is possible.

DAGs will be used throughout this thesis to describe assumptions of the study data structure and the methods employed to estimate causal effects of interest. Generally, it is important to keep in mind that DAGs are just as good as the background information used to draw them and that they make no indication to how likely or strong potential bias is. It is possible to represent the same research question with different DAGs, and it might be useful to repeat the analysis for different DAGs to improve understanding of possible causal effects. [26, 27]

1.6 Randomized and non-randomized studies

Two primary study types are commonly applied in epidemiology: randomized controlled trials (RCTs) and observational studies. [29] Perfect RCTs are considered to be the gold standard in medical research to quantify causal effects. Their key features are the comparison of an intervention (e.g. a treatment) of interest with a control, randomized assignment of individuals to the treatment and control group, and a form of blinding to hide which treatment has been given, in order to limit the possibility of bias arising, from participants or those administering the treatment. [1, 2, 3] RCTs prospectively follow the effects of a treatment on individuals from a well defined starting point forward in time. In its basic form, the parallel group design RCT, individuals are assigned into the treatment group ($\mathsf{X} = 1$) and the control group ($\mathsf{X} = 0$). This assignments takes place randomly, meaning independent of individuals' characteristics to ensure balanced characteristics between groups and the fulfillment of the exchangeability assumption. If the treatment assignment is perfectly predicted by randomization, no other variable can have an influence on X and the causal effect can be estimated consistently without confounding bias. [3] Figure 1.4 depicts the DAG for a perfect RCT with R indicating the randomization decision.

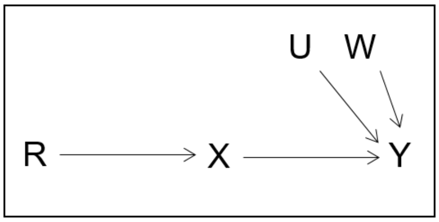


Figure 1.4: Visualization of a perfect RCT in which R*: randomization perfectly predicts* X*: treatment group assignment. Hence, no other confounders (*$\mathbf{W}$ *or* $\mathbf{U}$*) can influence* X*. This means that there are no arrows from either* U *or* W *into* X*, which makes it straightforward to estimate the causal effect* X → Y *from their association.*

Perfect RCTs make the assumption of perfect compliance to assigned treatment, lack of selection bias due to differential follow-up in the course of the trial, and no confounding after randomization. By design they fulfil the exchangeability assumption and therefore make consistent estimation of the causal treatment effect possible. [3, 13] If treatment and control protocol are not followed exactly, the RCT is said to be imperfect. Non-compliance exists when individuals do not receive the treatment they have been randomized to. It occurs after treatment assignment an therefore can be affected by participants' characteristics. In this case a 'per protocol analysis' in which all non-compliant participants are excluded, could lead to biased treatment effect estimation. Often an 'intention to treat' (ITT) analysis is employed which estimates the effect of assigned treatment, regardless of the actual treatment received. [25, 30]

Using RCTs to investigate causal effects is not always possible. They might be not feasible if the exposure studied is not manipulable, or it is unethical or too expensive to do so. Furthermore, if they are possible, RCTs may need to employ strict exclusion criteria for participation to ensure the safety of participants. This might lead to the exclusion of important subgroups, such as older, frailer patients or pregnant women, which will diminish the generalisability of the study results and reduce possibility for subgroup analyses. [3, 4, 29]

Observational studies are often employed to overcome these limitations and study important health research questions in broad patient populations that could not have been addressed by RCTs. Examples of observational data sources are surveys on healthcare providers, censuses, insurance claims records, administrative records or patients records from healthcare facilities. [4, 5] Many different observational study designs are possible, and further detailed explanations can be found elsewhere, for example in Lu [31] or Thiese [32]. In this thesis we focus on prospective cohort studies for which the participants are observed to have received either the treatment of interest or the control treatment, and are then followed up for a given period, at which point their outcome status is measured. A key feature of all observational study designs is that the exposure is not randomly assigned, but often treatment choice/ exposure depends on individuals' characteristics. Due to the lack of randomization of treatment assignment, characteristics between the treatment groups might not be balanced and the exchangeability assumption not fulfilled. If these characteristics also affect the outcome of interest, estimation of the causal effect of treatment will be influenced by confounding bias. The presence of confounding and how to account for it is therefore a major concern for the causal analysis with observational data. As the design of observational studies cannot rule out the risk of confounding, statistical methods which aim to adjust for confounding are employed for the estimation of the treatment effect. In the following, methods which aim to account for measured and unmeasured confounders are reviewed. These methods are studied in depth in specific chapters of this thesis.

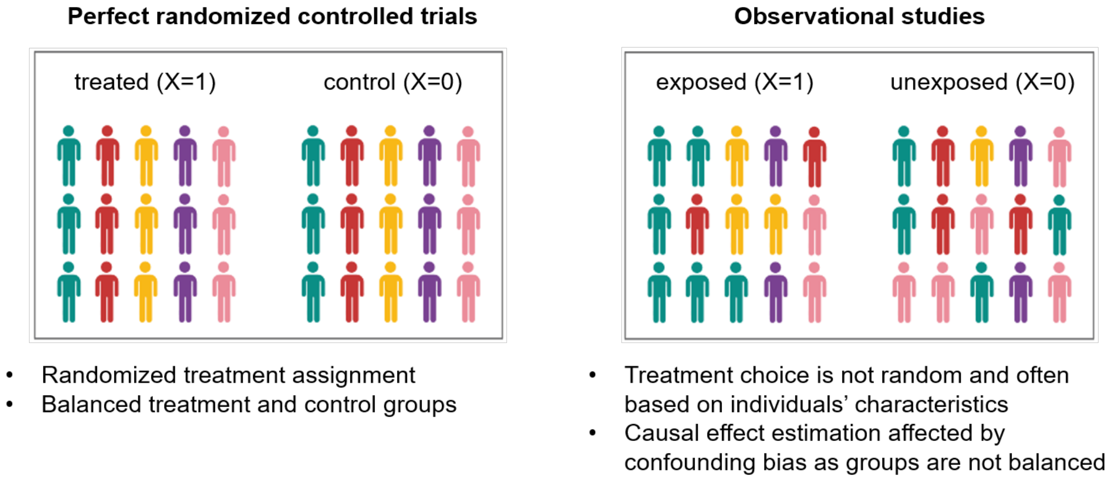

Perfect randomized controlled trials
treated (X=1)
control (X=0)
Randomized treatment assignment
Balanced treatment and control groups
Observational studies
exposed (X=1)
unexposed (X=0)
Treatment choice is not random and often based on individuals' characteristics
Causal effect estimation affected by confounding bias as groups are not balanced

Figure 1.5: Summary of key differences between perfect RCTs and observational studies.

In Figure 1.5 the main difference between RCTs and observational studies is visualized, which is the assignment to the treatment groups and its consequences for the comparability of treatment groups with regard to participants' characteristics. These are represented by icon colours.

1.7 Evidence-based medicine and triangulation of causal evidence

Evidence-based medicine (EBM) describes the integration of best research evidence for a particular question with clinical expertise and patient values. [33] Evidence to guide treatment decisions in clinical practice can be derived from various data sources and research designs with different strengths and limitations, such as RCTs and observational studies. The evidence pyramid is often used to order existing research designs into a hierarchy of informative value. One example of the evidence pyramid is depicted in Figure 1.6 similar to the pyramid provided by Rosner [34].

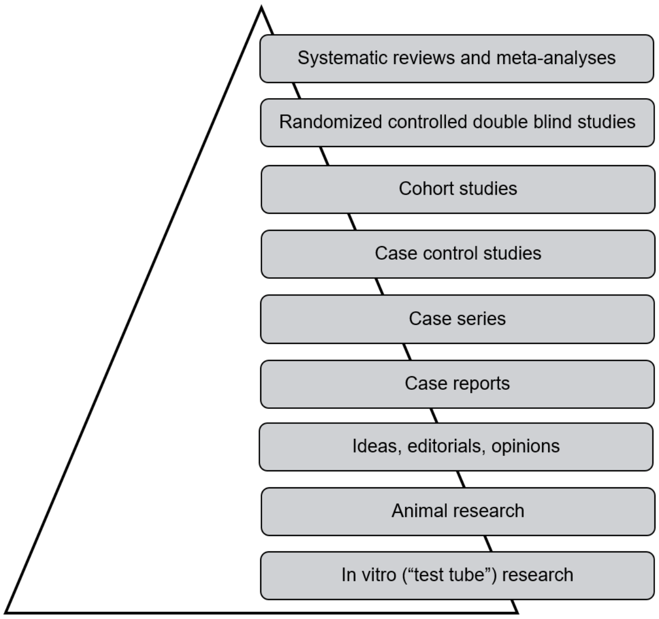

Systematic reviews and meta-analyses
Randomized controlled double blind studies
Cohort studies
Case control studies
Case series
Case reports
Ideas, editorials, opinions
Animal research
In vitro ("test tube") research

Figure 1.6: Example of the evidence pyramid of evidence-based medicine. A similar pyramid is provided by Rosner [34].

In this hierarchy, evidence from research designs with higher risk of bias are grouped at the bottom of the pyramid, such as in vitro research, animal research or editorial/ expert opinions. Higher valued evidence sources regarding rigor and freedom of bias are organised at the top of the hierarchy. In references to these aspects, RCTs are regarded the highest valued source to generate reliable and unbiased evidence from. Nevertheless, EBM advocates for pluralistic inquiry of research questions via the integration of evidence from different sources on the hierarchy. [35, 36, 37] As explained above, RCTs have limitations which affect their feasibility and generalisability for certain research questions. Non-randomized observational studies are an alternative where RCTs are not possible. Their value of evidence relies on three key principles: (1) a well-defined target population, (2) analysis and reporting that includes study results of patients of the cohort and (3) application of appropriate methodology and statistics regarding data structure and nature of the research question. [38]

Similar to the idea of integrating evidence from different research designs, the concept of 'triangulation' refers to the strategic use and comparison of multiple analytical approaches to address one research question. [8, 39] This concept is applied to strengthen the robustness and transparency of causal analysis. As explained above, causal analysis relies on often untestable assumptions about data structure and potential sources of confounding. [8, 40] The analysis can be strengthened and reliability of the results can be increased by employing different statistical methods and thereby looking at the research question at hand from different angles. These methods must all estimate the same underlying causal effect and ideally are subject to different and unrelated sources of bias. Triangulation is closely related to the concept of consistency of research results. If methods for triangulation are chosen with differing directions of bias, similar estimation results of the causal effect would give confidence in the evidence reliability. [8, 9, 10] Triangulation is a prospective approach and is most powerful if fundamentally different methods are chosen for comparison as to ensure that their sources of bias are different, and operate in different directions. This makes it necessary for researcher to consider and communicate assumptions and limitations of each method explicitly and will also foster collaboration between researchers working with different analytical methods. [8] Although triangulation can be done in a purely qualitative manner by comparing estimates from different data sources and study designs, ideas have been developed for the quantitative synthesis of estimation results within the same study, for example the 'Triangulation WIthin a STudy' (TWIST) framework [11].

1.8 Methods for estimating causal effects in observational data

Different analytical methods exists to estimate causal effects from observational data. They rely on different sets of assumptions and differ in their ability to adjust for measured and unmeasured confounding. The focus of this thesis lies on the development and application of a causal treatment effect estimation framework to derive suitable evidence from observational data and answer important questions about the safety and effectiveness of type 2 diabetes treatments. Triangulation of evidence from different analytical methods is a crucial part of this framework in order to strengthen the causal analyses. In the following, relevant statistical methods used for controlling measured and unmeasured confounding in observational research are summarized. The main emphasize will lie on methods addressing unmeasured confounding, with particular focus given to the Instrumental Variable method.

1.8.1 Notation

Before introducing relevant statistical methods to estimate the treatment effect of interest, the necessary mathematical notation is summarized in the following. This notation will be used throughout this thesis.

We are interested in estimating the causal treatment effect β of the treatment decision on the outcome of interest. Assume a study population of size N is analysed, which is itself clustered into J disjoint sets representing treatment decision making healthcare providers. Provider j treats n_j patients, so that $N = \sum_{j=1}^{J} n_j$. Within each provider, the patients' index, $i = 1, \dots, n_j$, is assumed to coincide with the order in which they have been treated, from first to most recent. The outcome of interest for patient i of provider j is denoted by Y_{ji}. Likewise, binary treatment variable X_{ji} denotes whether a patient receives treatment A ($X_{ji} = 0$) or treatment B ($X_{ji} = 1$). Measured and unmeasured confounders are represented by the G- and M-length vectors $\mathbf{W_{ji}}$ = ($W_{1ji}, \dots, W_{Gji}$) and $\mathbf{U_{ji}}$ = ($U_{1ji}, \dots, U_{Mji}$), respectively. For the population $\mathbf{W}$ and $\mathbf{U}$ are matrices of the size $N \times G$ and $N \times M$. In the following, the indexes i and j will be omitted from the notation if explanations are not specific to certain providers or individuals. Subpopulation sizes for the individuals treated with $X = 1$ and $X = 0$ are N_{Tx} and N_{Ct} respectively.

1.8.2 Regression adjustment and Propensity score matching

The causal effect β can be estimated consistently by adjusting for all measured confounders in a multivariable regression, under the NUC assumption. [20] This estimate is denoted as the 'Corrected as Treated' (CaT) estimate in the following and is defined as:

$$\hat{\beta}_{CaT} = \hat{E}[Y|X = 1, \mathbf{W}] - \hat{E}[Y|X = 0, \mathbf{W}]$$

and estimated from fitting the multivariable regression

$$E[Y|X, \mathbf{W}] = \beta_{Y,0} + \beta_{CaT}X + \beta_{\mathbf{Y},\mathbf{W}}\mathbf{W}, \qquad (1.1)$$

where $\mathbf{W}$ is a $N \times G$ matrix that forms a sufficient adjustment set (and therefore satisfying the NUC assumption) and $\beta_{\mathbf{Y},\mathbf{W}}$ is a $G \times 1$ column vector.

Another analytical approach to account for systematic differences between the treatment groups due to the lack of randomization in observational data is the propensity score matching (PSM) method. The propensity score is a balancing score and is defined as the probability of an individual to receive the treatment conditional on all observed covariates [41, 42]:

$$\pi_i = Pr[X_i = 1|\mathbf{W}].$$

The true propensity score is not known in observational studies but can be estimated using the data at hand for example by fitting a logistic regression model:

$$Logit(Pr[X = 1|\mathbf{W}]) = \beta_{X,0} + \beta_{\mathbf{X},\mathbf{W}}\mathbf{W},$$

where $\mathbf{W}$ is a $N \times G$ matrix and $\beta_{\mathbf{X},\mathbf{W}}$ is a $G \times 1$ column vector. Hence, the estimated propensity score $\hat{\pi}_i$ is the predicted probability to receive the treatment derived from fitting this model. With this definition of π_i and in a set of individuals with the same propensity score, under the NUC assumption the distribution of measured confounders will be the same between treatment groups. [42] For example, propensity score matching makes it possible to estimate the ATE by forming matched sets of treated and untreated individuals with similar propensity scores. [41] A summary and explanation of different matching strategies can be found for example in Austin [42]. In the following studies '1-1 matching' was applied for which one untreated individual is matched to one treated individual. In order to find the most similar propensity score for matching 'nearest neighbour matching' is applied, which matches untreated individuals to a treated individual with the closest propensity score. In case of more than one identical closest propensity score, an untreated individual is chosen randomly from this subset. After this matching process, the treatment effect can be estimated from the matched data. In order to reduce residual differences in measured baseline characteristics, regression adjustment can be applied. [42], Therefore, the target treatment effect can be estimated from the matched sample with the regression model, similar to model (1.1):

$$\mathsf{E}[\mathsf{Y}^{\circ}|\mathsf{X}^{\circ}, \mathbf{W}^{\circ}] = \beta_{\mathsf{Y},0} + \beta_{\mathsf{PSM}}\mathsf{X}^{\circ} + \beta_{\mathbf{Y},\mathbf{W}}\mathbf{W}^{\circ},$$

where the symbol $^{\circ}$ indicates that this model is estimated in the matched data only, $\mathbf{W}^{\circ}$ is a $\mathsf{N} \times \mathsf{G}$ matrix. Regression adjustment on the matched data can help increase the precision of treatment effect estimates for continuous outcome and increase statistical power for continuous, binary and time-to-event outcomes. [42, 43, 44] A limitation of the approach is that the 1-1 nearest neighbour matching must be applied on a complete case dataset for all measured confounders. Estimation of π_{i} and β_{PSM} might be therefore biased in case missingness is not missing completely at random. [45] Additionally, it is possible with this matching strategy, that not all individuals can get matched which will lead to a loss of information. It is also important to reiterate that the estimation of the treatment effect using propensity score matching is done under the NUC assumption, hence the method is not able to account for unmeasured confounding. [46] Other related methods which make use of π_{i} are stratification on the propensity score, inverse probability on treatment weighting and covariate adjustment on the propensity score. These approaches were not employed in the following studies, but further explanations can be found for example in Austin [42], Rosenbaum and Rubin [41] or Austin and Mamdani [47].

1.8.3 The prior event rate ratio and difference-in-difference approach

Standard regression adjustment and PSM utilize data on confounders measured at baseline/ initiation of the study treatment. It is also possible to adjust for events occurring before the study treatment, such as experiences of adverse event and comorbidities. Both approaches work under the NUC assumption and are therefore contingent on measured confounders forming a sufficient adjustment set. The prior event rate ratio (PERR) approach is applied specifically to time-to-event data from more than one distinct time period to overcome bias due to unmeasured confounding. It is a suitable approach to estimate causal effects on repeatable, non terminal outcomes, for example adverse effects such as infections. The outcome modelled in the prior period and the study period using Cox proportional hazard models. [48, 49, 50] A visualization of the before-and-after design or self-controlled design [51] of the PERR method, similar to Lin et al. [49] and Rodgers et al. [50] is shown in Figure 1.7.

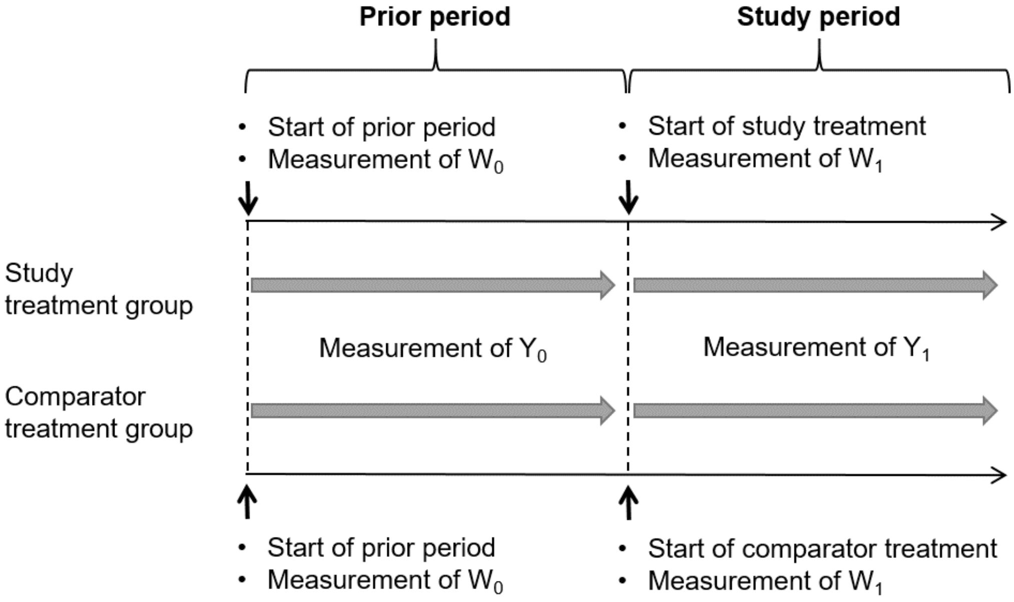


Figure 1.7: Before-and-after design of the PERR approach utilizing data from the prior and study period, similar to the visualization of Lin et al. [49] and Rodgers et al. [50]

The prior period starts before and ends with the study treatment initiation. It is important that neither of the two treatment groups have been exposed to the study treatment in this period. The study period starts with the initiation of the study treatment of interest or its comparator treatment. The outcome of interest as well as relevant confounders are measured in both periods. Let Y_0 and Y_1 denote the time-to-event outcome for the prior period and the study period respectively. Similarly, let W_0 and W_1 denote the relevant measured confounders in both periods. The follow-up time in prior and study period expands until the end of follow-up data of the respective period, occurrence of the outcome of interest or a censoring event. The idea of the PERR approach is to use the treatment effect measured for the prior period to capture the degree of unmeasured confounding and adjust for it in the estimation of the treatment effect of interest in the study period. It presumes that the estimated 'treatment effect' hazard ratio (HR) in the prior period reflects the effect of measured and unmeasured confounders on the outcome, but without any contribution from the actual treatment itself. By contrast, the study period estimated hazard ratio is assumed to be influenced by the true (causal) treatment effect, plus the same magnitude of confounding bias as the prior period. By taking the ratio of the prior and study period hazard ratio estimates, the PERR estimate:

$$\widehat{\mathsf{HR}}_{\mathsf{PERR}} = \frac{\widehat{\mathsf{HR}}_{\mathsf{prior}}}{\widehat{\mathsf{HR}}_{\mathsf{study}}},$$

effectively removes confounding bias by cancellation, leaving only the true causal effect. [48] Typically, $\widehat{\mathsf{HR}}_{\mathsf{prior}}$ and $\widehat{\mathsf{HR}}_{\mathsf{study}}$ are the hazard ratios from fitted Cox proportional hazard models of the prior and study outcome regressed on W_0 and W_1 respectively. [50] Standard error estimates for the causal estimate are derived via bootstrapping. [48] Due to the non-linearity of Cox models, the PERR method has been shown to result in attenuated treatment effect estimates. The PERR-ALT approach and its extension PERR Pairwise have been developed to overcome this drawback. [49, 52] Important assumptions of the PERR approach in order to estimate the causal treatment effect consistently are

1. **Independence of treatment decision and prior outcome:** The treatment decision X needs to be independent of Y_0. This assumption would be violated if the outcome of interest in the prior period affects the treatment decision of the study period for example in case of studied adverse effects.

2. **Time-invariant unmeasured confounding:** This assumption requires the

effect of U on the outcome Y_0 and Y_1 to be constant across time conditional on W_0 and W_1 respectively.

It has been shown in simulation studies that violation of these assumptions can lead to biased treatment effect estimates. [52, 53] Further investigations of assumption violations are outlined in Chapter 3. Due to this possibility of bias, it is important to apply subject matter knowledge to justify the assumptions. [54] Furthermore, it is important to chose the length of both periods with care. It needs to be long enough to be able to detect Y_0 and Y_1, while keeping in mind that the influence of U on the outcomes remains constant for both periods. [50]

For continuous outcomes the difference-in-difference (DiD) approach is analogous to the PERR approach. The treatment effect in the study period is corrected using the treatment estimate of the prior period. This is operationalized in a regression model for the study outcome, aggregated at treatment group level with a period-treatment interaction. [55, 56] The causal effect is estimated with the following model:

$$\mathsf{E}[\mathsf{Y}^*|\mathsf{X}^*, \mathsf{W}^*, \mathsf{P}^*] = \gamma_0 + \gamma_\mathsf{P}\mathsf{P}^* + \gamma_{\mathsf{X}^*}\mathsf{X}^* + \beta_{\mathsf{DiD}}\mathsf{P}^* \cdot \mathsf{X}^* + \gamma_\mathsf{w}\mathsf{W}^* + \gamma_{\mathsf{WP}}\mathsf{W}^* \cdot \mathsf{P}^*.$$

Here, $\mathsf{X}^* \in \{0, 1\}$ is a 2N-length treatment indicator variable, $P^* \in \{0, 1\}$ is the period indicator variable of the same length with $\mathsf{P}^* = 0$ indicating the prior period and $\mathsf{P}^* = 1$ the study period. The variables $\mathsf{Y}^* = (\mathsf{Y}_0, \mathsf{Y}_1)^\mathsf{T}$, $\mathsf{W}^* = (\mathsf{W}_0, \mathsf{W}_1)^\mathsf{T}$ summarize the information of outcomes and measured confounders for both periods in a vector of the same size. The DiD estimate of the causal treatment effect is the regression coefficient of the $\mathsf{P}^* \cdot \mathsf{X}^*$ interaction term. [57] Fitting this model also facilitates the estimation of standard errors, as they can be taken directly from the hessian matrix.

In Chapter 3 the DiD regression approach will be utilized to estimate the relative risk an adverse effect for two oral type 2 diabetes treatments. Additionally, a formal proof of the assumptions outlined above is given in the Appendix of this Chapter.

1.8.4 The Instrumental Variable approach

The instrumental Variable (IV) approach addresses measured and unmeasured confounding in observational data by attempting to create a pseudo-randomized controlled trial. It does this by utilizing an IV, denoted with Z, which introduces random variation in the treatment X. [58, 59] In order for Z to be a suitable or 'valid' instrument, three important assumptions about its relationship with the treatment decision and outcome of interest must be satisfied. Additionally, Z will be regarded as a binary variable in this thesis. The DAG in Figure 1.8 summarized the assumed data structure.

1. **Relevance assumption:** The instrument Z needs to be strongly associated with X, as shown in the DAG of Figure 1.8 with the path Z → X.

2. **Exclusion restriction:** This assumption requires Z to be independent of Y given X, $\mathbf{W}$ and $\mathbf{U}$ and therefore affects the outcome only through the treatment decision. A direct path Z → Y is therefore shown in red and crossed out in the DAG.

3. **Exchangeability assumption:** The instrument Z does not share a common cause with Y, as depicted by the red and crossed out $\mathbf{U}$ → Z path in the DAG.

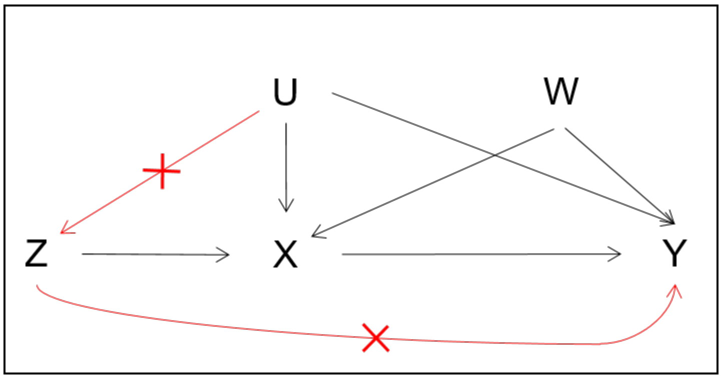


Figure 1.8: DAG depicting the assumed data structure of the Instrumental Variable method, including the variables X*: treatment decision,* Y*: outcome of interest,* Z*: Instrumental Variable,* $\mathbf{W}$*: measured confounders,* $\mathbf{U}$*: unmeasured confounders.*

For the simple case of a binary instrument, the IV estimate for the ATE of X on Y is given by

$$\beta_{\mathrm{IV}} = \frac{\mathrm{E}[\mathrm{Y}|\mathrm{Z}=1] - \mathrm{E}[\mathrm{Y}|\mathrm{Z}=0]}{\mathrm{E}[\mathrm{X}|\mathrm{Z}=1] - \mathrm{E}[\mathrm{X}|\mathrm{Z}=0]}. \tag{1.2}$$

[60, 61] That is, the ATE is estimated as the ratio of, the effect of Z on Y and the effect of Z on X. The idea of the IV method can be related to a perfect RCT as shown in Figure 1.4, where Z is replaced with the randomization indicator, R. Here, R also satisfies the three IV conditions described. Under complete adherence to randomisation, R predicts X perfectly and hence no other factors such as $\mathbf{W}$ or $\mathbf{U}$ can. In this case, the ATE would simply equal the effect derived from an ITT analysis. In the IV context, this is equivalent to the numerator in formula (1.2), since the demoninator is equal to 1. Returning to the RCT, if non-compliance to the treatment protocol is observed, so that R does not perfectly predict X, the ITT estimate will be a diluted or attenuated version of the ATE. The IV estimate in (1.2) overcomes this by dividing the ITT estimate by a denominator that lies somewhere between 0 and 1, and whose value decreases as the relationship between Z and X gets weaker. This inflates the ITT estimate so that it remains consistent for the ATE. For example, if half the patients comply and the denominator is equal to 1/2, the IV estimate for the ATE scales the ITT estimate up by a factor of 2. [58]

For the estimation of the ATE from observational data using IV methods, the three assumptions explained above are only sufficient to obtain estimates for the lower and upper limits of the causal treatment effect. It is possible to test the causal Null hypothesis as we are able to estimate the ITT estimate, the effect of Z on Y. If the instrument is valid, it will only have an effect on the outcome through X. Therefore, the effect of Z on Y should be zero if there is no causal relationship between X and Y. [58, 62] A forth assumption is needed in order to identify a point estimate β for the causal effect. Different variations of this fourth assumption have been defined in the literature and their distinction is important for the interpretation of the estimated effect. [63, 64]

4c. **Constant treatment effect:** This assumption requires identical (i.e. constant) treatment effects for all patients in the population, which is generally impossible for dichotomous outcomes and very unlikely for other types of outcomes. Hence, other less strict formulations have been proposed. [63]

4h. **Homogeneity:** Under this assumption the ATE in the population can be identified. It requires that no additive effect modification exists across the levels of the instrument Z and within the groups of exposed and unexposed.

[63, 64]

4m. **Monotonicity:** For this assumption to be fulfilled, the instrument needs to only affect X in one direction for all patients. In other words, it requires that the level of X for each patient given Z is a monotonic increasing function of the level of Z. [65]

The monotonicity assumption is in many studies more plausible and allows for the estimation of the local average treatment effect (LATE). The LATE is the ATE in the subgroup of compliers, which are individuals who are treated in accordance with the level of their IV. [58, 63, 64] In case of binary Z it is possible to create four compliance groups which are summarized in Table 1.1 for completeness. It is noteworthy that the group of compliers cannot be identified from the data at hand of an observational analysis, hence the relevance and practical translation of the LATE has been questioned. [66] For a binary Z the monotonicity assumption requires the absence of defiers in the study population. [63, 67]

Compliance group	Explanation
Always-taker	Individuals who receive $X_{ji} = 1$ regardless of Z_{ji}
Never-taker	Individuals who receive $X_{ji} = 0$ regardless of Z_{ji}
Compliers	Individuals who receive $X_{ji} = 1$ in accordance to Z_{ji}, with $X_{ji}^{Z_{ji}=1} = 1$ and $X_{ji}^{Z_{ji}=0} = 0$
Defiers	Individuals who receive $X_{ji} = 1$ in discordance to Z_{ji}, with $X_{ji}^{Z_{ji}=1} = 0$ and $X_{ji}^{Z_{ji}=0} = 1$

Table 1.1: Overview of compliance groups with respect to values of Z *and the treatment decision* X*. Here, both variables are binary.*

A common and general procedure to estimate β_{IV} in the case of a continuous outcome is the so called 'Two-Stage Least Squares' (TSLS) approach. [68, 69]. In the case of a binary treatment, the first stage usually comprises a logistic regression model for X adjusting for all measured confounders and the instrument:

$$\text{Logit}(\Pr[X = 1|Z, \mathbf{W}]) = \beta_{X,0} + \beta_{X,Z}Z + \beta_{\mathbf{X},\mathbf{W}}\mathbf{W}. \tag{1.3}$$

The fitted first stage model is used to predict X given Z and $\mathbf{W}$ as $\hat{X}$. The second stage model is the outcome model regressing Y (here: continuous) on the predicted values $\hat{X}$ and the measured confounders:

$$\mathrm{E}[\mathrm{Y}|\hat{\mathrm{X}}, \mathbf{W}] = \beta_{\mathrm{Y},0} + \beta_{\mathrm{IV}}\hat{\mathrm{X}} + \beta_{\mathrm{Y,W}}\mathbf{W}. \tag{1.4}$$

The TSLS estimate for the causal treatment effect β_{IV} is the coefficient of $\hat{X}$ from this fitted regression model. [68, 70]. Using a valid IV and given model (1.4) is correctly specified, the TSLS estimate is consistent under the homogeneity assumption. [71, 72].

An alternative estimation procedure for binary and rare outcomes and binary exposures is the ‘Control Function’ (CF) approach. [73, 74] The approach, as explained below is recommended as it is robust to heterogeneity in the magnitude of selection bias with respect to the instrument. [75] For the first stage of the CF approach the treatment decision model (1.3) is estimated. From the fitted model the residual $\hat{\Delta} = \mathrm{X} - \hat{\mathrm{X}}$ is calculated and used in the second stage model with

$$\mathrm{Logit}(\mathrm{Pr}[\mathrm{Y} = 1|\mathrm{X}, \mathrm{Z}, \mathbf{W}, \hat{\Delta}]) = \beta_{\mathrm{Y},0} + \beta_{\mathbf{Y},\mathbf{W}}\mathbf{W} + \beta_{\mathrm{CF}}\mathrm{X} + (\beta_{\mathrm{Y},\hat{\Delta}} + \beta_{\mathrm{Y},\mathrm{Z}\hat{\Delta}}\mathrm{Z})\hat{\Delta}.$$

For this estimation procedure the treatment effect is taken as the coefficient of X. This estimation procedure is showcased in Chapter 3, when triangulating results of different estimation methods for the relative risk of experiencing an adverse effect comparing two commonly prescribed oral type 2 diabetes treatments. In Chapter 5, causal effects for continuous treatment outcomes and binary adverse effects were estimated using the TSLS approach only. This is justified, as the results of TSLS and CF did not differ substantially in the triangulation analysis of Chapter 3 and with the aim to make this applied study more accessible for the target audience of clinicians treating people with type 2 diabetes.

Careful consideration is needed in order to find a valid instrument which meets the above explained assumptions (especially the first three assumptions) and is measured in the data at hand. A selection of possible IVs that have been proposed in the literature are: geographic information (e.g. distance to the closest hospital), time-based characteristics (date of therapy initiation) or historical/lagged variables (previous therapy as prescribing preference of healthcare provider). The method of Mendelian randomization uses genetic variants (e.g. single nucleotide polymorphisms) as instruments. [55, 76] In consecutive chapters, all IV analyses will be conducted using a healthcare provider prescription preference-based instrument. Therefore, in Section 1.8.5 important aspects for the use of this instrument and the framework under which it can be applied are discussed in more detail.

An important limitation of the IV method is its reliance on assumptions which are not all testable with the data at hand. Only the relevance assumption can be fully tested with the first model of the TSLS and CF procedure. The F-statistic for the coefficient of Z in this treatment decision model gives an indication on how strong Z is related to X. As a rule of thumb, Z is considered a strong instrument if the value of this F-statistic exceeds 10. [77] Weak instruments can lead to so called 'weak instrument bias' in the estimation of the causal treatment effect. Furthermore, it can lead to an amplification of bias related to other violation assumptions. [58] The exclusion restriction and exchangeability assumption can not be tested with the data at hand, as this would require knowledge about unmeasured confounders. They can only be falsified with the data at hand and subject matter knowledge. Additionally, evidence from previous studies can be applied to decide whether these assumptions are realistic for the research question and data at hand. [58, 63, 64, 70, 78] Violations of these two assumptions lead to bias in the estimation of the causal effect. [70] One way to partially verify the exchangeability assumption using the measured covariates is to tabulate their distribution across the levels of the proposed instrument. If the measured confounders appear to be unbalanced when grouped by instrument level, it is likely that Z is also confounded by U. In this way, the measured confounders are used as proxies for unmeasured confounders to assess the validity of the IV assumption. However, unmeasured confounding in observational studies can generally never be fully ruled out. [64, 70, 79]

1.8.5 Healthcare provider prescription preference as Instrumental Variable

Healthcare provider prescription preference (PP) has been proposed as Instrumental Variable, to exploit variation in prescribing behaviour which cannot be explained purely by patient demographics that are prognostic for disease, or regional variation in treatment guidelines. [80, 81] According to the systematic review by Widding-Havneraas et al. [59], PP has been increasingly used as Instrumental Variable in medical research, especially for treatment effect studies for cancer, cardiovascular diseases and mental health. Strong prescription preferences can be formed for example due to controversial treatment options or if providers are specialists for a specific treatment (e.g. surgeries). [80] Usually, the choice of treatment will be determined by an interplay between many factors which include but are not limited to: the providers subjective preference at a time given their cumulative experience in practice; patient characteristics, including their age, sex, comorbidities and contra-indications of any existing medications; a patient's insurance that only covers one specific treatment; a strong preference coming from the patient given their own experience. [82] Depending on data availability, provider prescription preference can be defined at three different levels of healthcare provider: geographical level such as regions [83, 84], facility-level such as hospitals, clinic or general practice level [85, 86], and on level of individual physicians/ practitioners [87, 88].

The proxy design and construction methods of the instrument

The preference of a provider for one treatment over another is a continuous variable, to account for the strength of this belief. It is very difficult to measure and requires a distinct survey design. [67, 89, 90]. Therefore, retrospective cohort studies with focus on the assessment of treatment effects and adverse risk often do not incorporate data on the true PP. Usually, for IV studies using preference based instruments, PP is conceptualized with a surrogate variable. This has been called the 'proxy design', which is represented in Figure 1.9. [58, 59, 91] In this design, the true underlying provider preference PP is assumed to be a valid IV and is approximated with the provider's manifest and observed prescribing behaviour utilizing information on X. This proxy variable is used as IV and hence denoted with Z. In this design, Z does not have a causal relationship with X but fulfils all IV assumptions described above.

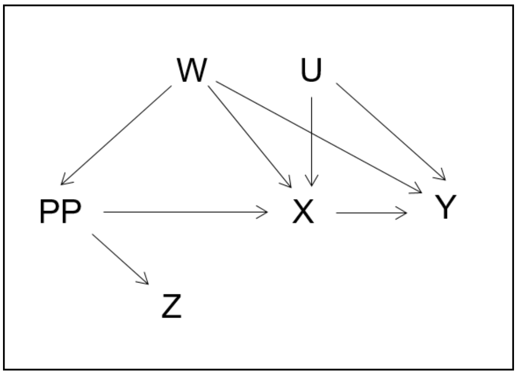


Figure 1.9: Description of the proxy design. Here PP: true unmeasured provider preference which is assumed to be a valid IV, Z: proxy variable of PP used as instrument, W: measured confounder, U: unmeasured confounder, X: treatment decision and Y: outcome of interest.

Different ways on how to construct a PP based IV with information on X have been proposed in the literature. In Chapter 4 the different construction methods are organised into two groups based on their use of the data at hand. Simple rule-based methods utilize data on treatment decisions only, to reflect on PP and construct the instrument Z. An often applied rule-based construction method was proposed by Brookhart et al. [81]. They suggested a binary IV that corresponds to the most recent prescription for each patient within a provider. Hence, $Z_{ji} = 1$ in case of $X_{ji-1} = 1$ indicates that provider j has a preference for treatment B at the prescription time of patient i. Furthermore, $Z_{ji} = 0$ in case of $X_{ji-1} = 0$ indicates a prescription preference for treatment A. Model-based construction methods are the second group of provider preference-based IVs. They use more complex statistical models to construct Z, taking the data structure of measured confounders into consideration. Further details and examples of rule-based and model-based construction methods are given in Chapter 4. As there are different construction methods available in the literature, Brookhart et al. [92] suggest to chose the method that appears to be most strongly related to the observed treatment decision, among those that are unrelated to measured confounders. An additional aspect that needs to be considered when choosing a suitable construction method is the fact that PP naturally evolves over time. This aspect should be ideally reflected in the construction of Z in order for this instrument to be a good proxy of PP and limit measurement errors. [92] Reasons for change in PP can be for example evolving experience of the provider, new efficacy and safety information from clinical trials, marketing efforts of pharmaceutical companies or a recent negative experience with the preferred drug. [92, 93, 94, 95, 96] Furthermore, the construction methods have different requirements on the data structure, such as for example the minimum amount of prescription information needed from each provider in order to calculate Z. Lastly, Ertefaie et al. [97] point out that non-ignorable missingness is a common occurrence in observational data and can lead to bias in the treatment effect estimation using a preference-based IV. Chapter 4 represents a comprehensive simulation study of different construction methods and their performance under the three mentioned data conditions: change in PP over time, provider size and missing data for measured confounders.

Assumptions of preference-based instruments

For the representation of a preference-based IV it has been assumed so far that PP is a valid instrument. This requires the IV assumptions as presented in Section 1.8.4 to be fulfilled. A PP-specific explanation of the assumptions are as follows:

1. **Relevance assumption:** Provider differ in their use of treatment independently of patient characteristics.

2. **Exclusion restriction:** A providers' use of the treatment is unrelated to the use of other medical inventions that might influence the outcome.

3. **Exchangeability assumption:** Patients are assigned to provider independently of the their prescription pattern.

4m **Monotonicity**: If a provider treats a patient with the study treatment, all providers with a preference equal or higher than the preference of this prescribing provider will also prescribe the study treatment. [92]

As a preference-based IV is most often used under the proxy design, Hernán and Robins [58] suggest a weaker version of the relevance assumption under which Z and X have to be associated either because Z causes X or because both variables share a common cause, such as PP. This assumption suggests that the treatment decision is partly made due to PP and not only based on guidelines or patient characteristics. [80] Therefore, prescribing patterns must still exist independently of patient characteristics. [88, 98] Boef et al. [90] conducted a study looking at survey data from physicians and 8 hypothetical patients, asking for their treatment decisions. Their study supports that provider preference at physician level is a determinant of treatment decision as only some variation of the treatment decisions was explained by physician characteristics or the patient population. This indicates that the relevance assumption is plausible for physician's treatment preference. [90] Ionescu-Ittu et al. [82] conduced a simulation study to assess the impact of weak preference-based IVs on the performance of the causal estimates. In this simulation study, strength of the IV is diminished by increasing the proportion of prescriptions that are independent of the providers preference. The results showed the application of a preference-based IV helped to remove bias due to unobserved confounding, but to a possible price of larger estimation variance. Whether this bias-variance tread-off favours the IV estimation approach compared to standard estimation approaches, will depend on the strength of the IV. [82]

The exclusion restriction will be violated if PP for one treatment is linked to the prescription of another treatment that also affects the outcome. For example, Lousdal [70] describes a study evaluating side effects for different treatments in which providers are found to prescribe nausea-relieving medications together with chemotherapy, as the studied treatment. In this case the preference-based IV will not be independent of Y as nausea is a side effect of chemotherapy. Another study by Newman et al. [99] assessing the treatment effect of phototherapy on hyperbilirubinemia in newborns discovered that the preference-based IV on hospital level was correlated to the treatment of infant formula. As infant formula is also a treatment for hyperbilirubinemia, the IV influenced the outcome through another path than the phototherapy use. Bias due to violation of this assumption can be accounted for by creating additional treatment categories related to the combination of treatments. Hence, IV methods can be used to estimate the effect of each treatment category. [92]

The exchangeability assumption will be violated if patients chose their healthcare provider based on the prescription preference, or if specialist provider with a preference for one drug would see sicker patients. Tabulating the distribution of the measured confounders grouped by the levels of the IV would reveal unbalanced groups. [64, 92, 100, 101] For the studies by Rassen et al. [100] and Bidulka et al. [83] it is assumed that a violation of this assumption in these two cases is unlikely or minimal if patients where seen by their usual primary care doctors. In their analysis, Davies et al. [88] assess this assumption by estimating risk differences of each measured confounding factor by patients' prescription X and by measures of their IV using robust standard errors clustered by provider. Additionally, the authors reported percentage reduction in the Mahalanobis distance, a summary measure of the total covariance imbalance, grouped by the levels of the IV and treatment decision. The authors found less covariate imbalance when grouped by the IV. [88] Davies et al. [91] propose to assess this assumption using negative control populations. A negative control population is a population that has a similar confounding structure than the study population but was not exposed to the treatment of interest. In a study using provider preference as IV, a negative control population can for example comprise patients that have been seen by the same healthcare provider who prescribed the treatment of interest to patients of the study population. The negative control patients were seen by the healthcare provider for an unrelated reason and therefore were not eligible to receive the treatment of interest. Is the IV associated to the outcome in the negative control population, this suggests that there is another mechanism with which the IV is related to the outcome other than through the treatment of interest. This could be for example an unmeasured common cause. [91]

Hernán [58] defined the monotonicity assumption for a continuous instrument using information on PP. The assumption requires that if a provider treats a patient with the study treatment, all provider with a preference for $X = 1$ equal or higher than the preference of this prescribing provider will also prescribe the study treatment. This assumption might be reasonable if patient characteristics can be combined into one single dimension, such as a propensity score, on which the provider bases the prescription decision. [67] But in reality, it is likely that provider use multiple patient characteristics for their treatment decisions and place different emphasis on these characteristics. Provider also might need to make treatment decisions against their preference due to patient characteristics such as risk factors and comorbidities. This could be an explanation why assumption 4m fails to hold as treatment decisions would not be ordered in alignment with the providers' prescription preference. [65, 67] Furthermore, as preference-based IVs are usually binary proxy instruments for the underlying and continuous prescribing preference, this assumptions is generally implausible. [58] Small et al. [65] distinguish this deterministic monotonicity assumption (assumption 4m(d)) as explained above from the less strict stochastic monotonicity assumption (assumption 4m(s)). It requires that the instrument needs to be related to X monotonically across subjects within strata of a sufficient set of measured and unmeasured confounders that are common causes of X and Y. The exact formulation of this assumption depends on the definition of the IV. [65, 90] It is generally not possible to verify monotonicity as we cannot observe which prescription an already treated patient would have been prescribed by another provider. Results from an IV analysis using preference-based IVs should be therefore interpreted with caution. [67] One possible way to assess this assumption is to survey healthcare provider about their treatment decisions on the same set of patients. This survey framework can make it possible to study the compliance group distribution in the target population and give insight into whether the treatment decisions are made in alignment with the monotonicity assumption. This assessment is specific to the study population and the definition of the instrument. Boef et al. [90] and Swanson et al. [67] conducted an assessment of assumption 4m using this survey framework and a preference-based IV. Boef et al. [90] used matrix plots visualizing the prescription pattern of the study treatment over providers and for each patient. The results showed that the prescription patterns were not consistent with assumption 4m(d). Their assessment of assumption 4m(s) was dependent on the specific construction methods of Z which utilized the treatment decision of the most recent patient. [92] The results showed that the probability of receiving the study treatment was higher or equal for all patients if the previous patient was also prescribed the study treatment. Form this, the authors concluded that the IV as defined by Brookhart et al. [92], was related in the same direction for all cases and the assumption 4m(s) was therefore not falsified. [90]

The above explanations and cited studies show, that the application of preference-based IVs relies on strict assumptions and the treatment effect results can be difficult to interpret. For studies using preference-based IVs the assumptions should be therefore assessed carefully using the proposed statistical methods and subject matter knowledge.

1.9 Conclusion

Evidence from observational studies has the potential to contribute towards answering cause-and-effect questions in medical research. Crucial for the quality of observational evidence is the use of appropriate methods for the estimation of the treatment effect of interest. Methods need to by employed that mitigate the risk of bias due to confounding. Regression models adjusting for measured confounders and the propensity matching method are able to estimate the treatment effect under the no unmeasured confounding assumption. Other methods which additionally address unmeasured confounding are the difference-in-difference approach and the Instrumental Variable approach. All introduced methods rely on different parts of the data at hand and data structure assumptions. Potential bias in the treatment effect estimation might origin from violations of different assumptions. This makes the collection of estimation approaches interesting to be explored in an triangulation framework. This idea is developed further in Chapter 3.

Furthermore, the Instrumental Variable method utilizing variation in prescriber preferences requires additional considerations, as this instrument is rarely measured in observational studies. Instead a proxy variable for prescriber preference is often used, for which different construction methods are proposed in the literature. The choice of construction methods for the proxy variable should aim to reduce measurement error of the true underlying prescription preference and relies on data aspects, such as possible changes of preference over time, missing data or the amount of available prescription data per provider. A more in depth analysis of the Instrumental Variable approach based on the proxy design for provider prescription preferences and different data conditions is outlined in Chapter 4.

Chapter 2

Introduction to type 2 diabetes research

2.1 Introduction

This chapter provides an introduction to diabetes research and discusses important aspects of the management of type 2 diabetes (T2D) in clinical practice. The number of people with diabetes is rising worldwide and its treatment causes a high financial burden on healthcare systems. [102, 103, 104] The majority of patients live with T2D. [105] The treatment of T2D is complex and clinicians need to make individualized treatment decisions based on patient preferences and their circumstances for example regarding comorbidities. [106] One newer treatment class for T2D are Sodium-glucose Cotransporter-2 Inhibitors (SGLT2is). Agents of this class act by inhibiting the reabsorption of glucose in the kidney in case of excess plasma glucose concentration which leads to increased excretion of glucose in the urine. [107, 108, 109] They have been prescribed more often since their introduction in 2013. Furthermore, recent treatment guidelines changes have opened prescription of this class to a wide patient population, due to strong RCT evidence of cardiorenal benefits and low risk of hypoglycaemia. [110, 111] It is these benefits that make SGLT2is especially interesting in older adults, for example over 70 years, which is an important subpopulation encompassing the majority of T2D patients. [112]

Nevertheless, knowledge of the risk and benefit profile of SGLT2is for older adults is limited as clinical trials employ strict exclusion restrictions and do not represent the heterogeneous group of older T2D patients adequately. [113] Chapter 1 introduced the idea of integrating evidence from observational research in health research and the use of causal inference methods to mitigate the risk of confounding bias in the estimation of treatment effects. In this chapter, these ideas will be developed further to undermine how evidence from real-world data can help close the knowledge gap of current T2D treatment guidelines and give further insights in the risk and benefit profile of SGLT2is, especially in the important patient population of older adults.

Chapter 3 and Chapter 4 will outline application case studies of a causal inference framework analysing the risk of genital infections, a well known adverse event of SGLT2is. Chapter 5 represents a causal observational study on the safety and effectiveness of SGLT2i regarding important treatment outcomes and adverse effects that are potentially associated with SGLT2i and especially dangerous for older individuals with T2D. For all studies, a T2D cohort from the Clinical Practice Research Datalink is utilized. This real-world database of primary care records from UK general practitioners is explained in more detail in the following.

2.2 Epidemiology, pathophysiology, diagnosis and complications of diabetes

Diabetes is a complex, chronic and progressive metabolic disorder, mainly clinically characterized with increased blood glucose levels (hyperglycaemia) and with an inherent risk of micro- and macrovascular complications. [114] The number of diabetes cases worldwide is rising and it is considered to be one of the major global epidemics of the 21th century. [102] It was estimated that in 2021 537 million people were affected by this disorder with projected numbers rising rapidly to 783 million people by 2045. T2D makes up over 90% of diabetes cases world wide. [105] In the UK 3.9 million people live with T2D and 2.4 million people have an increased risk of T2D due to increased blood glucose levels. [115] Treatment of this disorder and its complications causes a high financial burden on healthcare systems. It is estimated for example that diabetes care accounts for 10% of the UK National Health Service (NHS) expenditures and will rise up to 17% by 2035/2036. [103, 104]

The cause of T2D is mainly unclear but its development is driven by two factors that affect metabolic balance: a defect in the insulin secretion by pancreatic β-cells, and insulin resistance when insulin-sensitive tissues are unable to respond to insulin effectively. [105, 116, 117] Type 1 diabetes (T1D) on the other hand is a chronic autoimmune disorder and is characterised by the destruction of β-cells and absolute insulin deficiency. [118] Insulin is an essential hormone needed to allow glucose from the blood stream to enter the body's cells in order for it to be converted into energy or stored. Insulin deficiency or resistance can lead to hyperglycaemia which can cause damage to peripheral tissues such as the cardiovascular and nervous system, kidneys, retina, and liver when left unchecked. Common potential complications include cardiovascular disease (CVD), diabetic retinopathy, nerve damage (diabetic neuropathy), kidney damage (diabetic nephropathy), and liver fibrosis as a late emerging complication. [105, 119] Known risk factors for the development of T2D are overweight and obesity, family history and genetic predisposition, ethnicity and age. [105, 117]

T2D is diagnosed in primary care in the UK with a blood test and the measurement of Haemoglobin A1c (HbA1c). This is a summary estimate of blood glucose levels over approximately 3 months. Diabetes is diagnosed in case of an HbA1c level of 48 mmol/mol (6.5%) or higher. In case of a diabetes diagnosis, T2D is diagnosed based on the exclusion of T1D and other rare forms of diabetes for example due to clinical features at presentation, additional autoantibody testing, or response to treatment. [120, 121] Besides hyperglycaemia, symptoms of undetected diabetes are for example polyuria, polydypsia and polyphagia, unexplained weight loss, frequent fatigue, and excessive thirst. As these symptoms often have gradual development they can remain unnoticed or might get disregarded, which might delay diagnosis. [105, 122]

2.3 Treatment of type 2 diabetes

The main focus of T2D management and treatment is the achievement and maintenance of adequate blood glucose control in order to avoid or delay diabetes specific complications. [106] UK, EU and US treatment guidelines recommend an HbA1c target of 58 mmol/mol (7.5%) for most individuals, but HbA1c targets should be individualized based on patient specific aspects such as individual's preferences, comorbidities, or risks from polypharmacy. [106, 111, 123, 124] Initial actions include patient education about beneficial life-style changes, such as exercise, weight management and dietary advice. [106] As first-line treatment to achieve blood glucose management, usually Metformin (MFN) is recommended, except in case of contraindication. If agreed on blood glucose targets are not achieved, healthcare provider have the option to intensify treatment. A detailed overview of non-insulin glucose-lowering oral medications are given in Davies et al. [106] including aspects such as primary mode of action, advantages and disadvantages such as adverse effects (AEs). In short, the major non-insulin oral medication classes for treatment intensification are Sulfonylureas (SU), Thiazolidinediones (TZD), glucagon-like peptide 1 agonists (GLP-1), Dipeptidyl peptidase-4 Inhibitors (DPP4i) , and Sodium-glucose Cotransporter-2 Inhibitors (SGLT2i). Treatment classes are comprised of several drugs, but will be referred to on class level if not stated otherwise. These treatment options differ in their mode of action, costs, safety profile and performance with respect to a range of health outcomes. [106]

International and UK guidelines have been updated recently and now give drugs within the SGLT2i class a prominent position in the treatment intensification pathway and as first-line treatment (if MFN is contraindicated). [110, 111] This update was made on the base of strong RCT evidence concluding that SGLT2is reduce the risk of chronic kidney disease (CKD) progression and CVD outcomes which was found to be independent of glycaemic control. [107, 108, 109] According to the guidelines, all patients with chronic heart failure or established atherosclerotic CVD, as well as patients with high risk of CVD, measured with a Cardiovascular Risk Score (QRISK2) of 10% or higher/ elevated lifetime risk, are eligible for treatment intensification with SGLT2i. [111, 124] This opens prescription of SGLT2i to a large T2D patient population. Indication of this treatment class is however avoided in cases where patients have increased risk of diabetic ketoacidosis (DKA) or CKD. CKD is indicated with the glomerular filtration rate (eGFR) of the kidney, and different thresholds are applied for different drugs within the SGLT2i class. In subsequent analyses outlined in this thesis a conservative threshold of eGFR $< 45\text{mL}/\text{min}/1.73\text{m}^2$ will be used as exclusion criteria to identify patients with CKD. If there is no clear indication for SGLT2i, or SGLT2i is contraindicated, other recommended second-line treatments are DPP4is, SUs and TZDs. [111, 125]

Since the introduction of SGLT2i (first approved in 2013) and the more recent guideline changes (update of NICE guidelines in 2022) SGLT2i prescriptions have increased. [106, 126, 127, 128] A recent cross-sectional study by Bidulka et al. [129] found that despite this increase, SGLT2i prescribing patterns are not yet following the guidelines regarding their suitability for patients with established atherosclerotic CVD. They found that people with prevalent CVD had a lower probability to be prescribed SGLT2i compared to patients without CVD. Barriers for the uptake of SGLT2i prescriptions from clinicians perspective are safety concerns and a lack of understanding of the benefit of SGLT2i, which could lead to patients not receiving SGLT2i even though they would benefit. [130] This stresses the importance of investigating the risk and benefit profile of SGLT2is thoroughly. Application studies presented in this thesis will therefore focus on important safety concerns and effectiveness analysis of this drug class.

2.4 Type 2 diabetes in older adults

Demographic changes worldwide mean that older people are the fastest growing segment of society. The rapid increase in obesity has in turn lead to a substantial increase in the rate of older people living with T2D. [112] In 2021, it was estimated that the diabetes prevalence was the highest in people aged 75–79 years. [131, 132, 133] Although not fully elucidated, many etiologic reasons are suspected to affect organ systems and tissues during the aging process, that will increase diabetes risk. [113] Examples of possible age-related factors are reduced insulin activity and deterioration of insulin secretory capacity, changes in body composition which favour the accumulation in adipose tissue, the increase use of medications which increase hyperglycemic propensity, or hormonal dysregulation and inflammation. [113, 134, 135, 136]

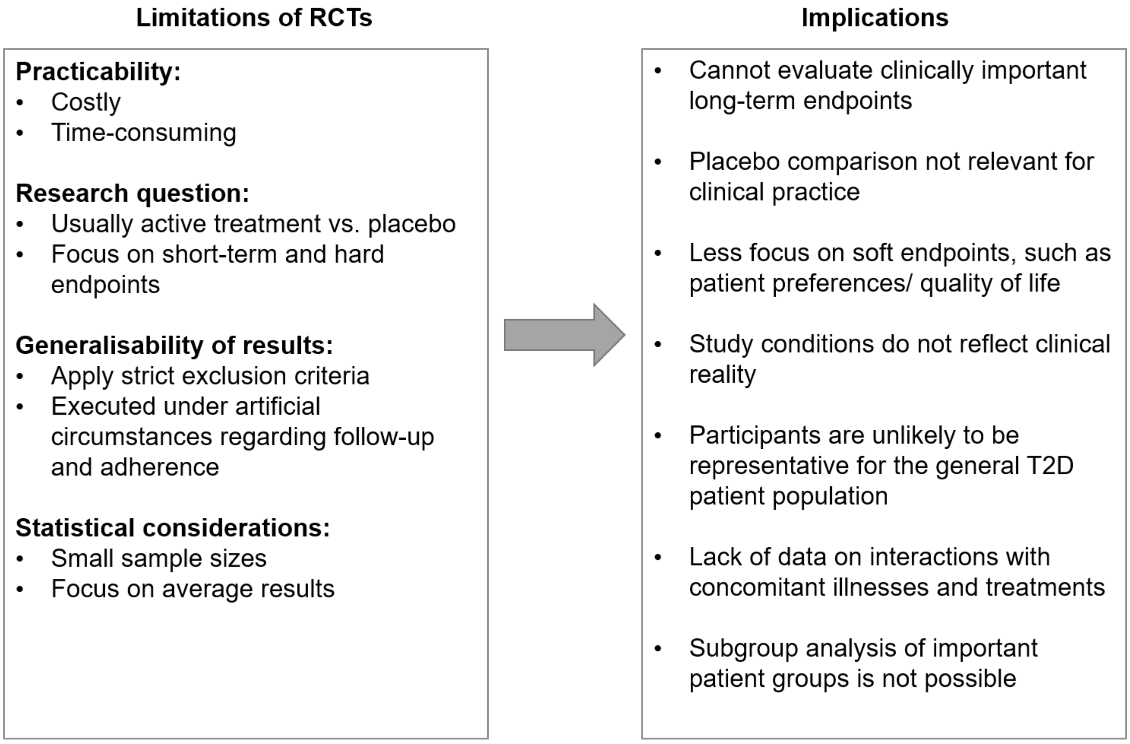
There is no universally agreed age-definition for the term ‘elderly’ or older adults. One distinction made is between ‘young-old’, 65-80 years old and ‘old-old’, 80 years of age or older. Even though this distinction is too simplistic to base clinical decisions on. [113, 131] Often, the somewhat arbitrary criteria of the age of 60 or 65 years has been applied as a cut-off for older adults, referencing the age of retirement in many developed countries. [137] Furthermore, this group of T2D patients have been described as heterogeneous as they present with a broad range of characteristics regarding comorbidities, functional abilities, life expectancy, and socioeconomic status. [138] How diabetes presents in older patients is also heterogeneous and a major factor for distinction is the age at which the disease developed. Individuals diagnosed at middle-age show for example increased fasting hepatic glucose production and insulin resistance, as well as an abnormal insulin response to glucose load. Patients who were diagnosed with diabetes later in life often have normal hepatic glucose production and older, lean patients are less insulin resistant. [131, 139]

Treatment of T2D in older patients is complex due this heterogeneity and treatment decisions need to include considerations about life expectancy, functional status, age-related changes in physiology (e.g. Sarcopenia) and in pharmacodynamics. Also, varying degrees of treatment adherence in this patient group complicate T2D management, for example due to cognitive impairment, social isolation, depression, or anxiety. [133, 140, 141] For this patient group it is crucial to individualize treatment decisions and focus on the avoidance of hypoglycaemia as well as adverse effects. Simpler and safer treatment regimens should be preferred to lower the burden of care and achieve better tolerance. [113, 133, 142, 141]

Clinicians treating older T2D patients face a lack of guidelines for this specific group as guidelines mostly rely on evidence from RCTs. But clinical trials in older adults are sparse as individuals over the age of 65 are often excluded due to comorbidities. Also, trials are not designed to take the diverse functional status in this patient group into account. [113, 140, 143] Hence, majority of the clinical trial data establishing risk profiles, glycaemic targets, and therapeutic interventions for people with T2D are not applicable for large segments of the older patient population. In light of the high heterogeneity of this patient group and the lack of guidelines for individualized treatment decisions, there is a need for strong evidence-based data to find ways for treating older individuals with T2D. [113] Observational studies including older patients can provide insights that are needed to close this knowledge gap. Strengths of this study type have been explained in Chapter 1 and the following section focuses on an explanation of how real-world data can provide valuable evidence for T2D treatment guidelines.

2.5 The place of real-world evidence in treatment guidelines

Optimal prescribing and the choice of intensification treatment for T2D management requires a full understanding of the agents risks and benefits. Evidence-based treatment guidelines are an important tool for clinical practice that supports informed decision making. Traditionally, these guidelines are based on evidence about the efficacy and safety from RCTs. This evidence is especially important for newer drug classes such as SGLT2i and DPP4i. [114, 144] But, as introduced in Chapter 1, RCTs have drawbacks that limit their application to answer certain issues clinicians face in their real-life practice and their generalisability to a wider patient population. In Figure 2.1 a summary of important limitations of RCTs and their consequences is given with a focus of related issues to diabetes research. [145, 146, 147, 148, 149]



Limitations of RCTs

Practicability:
- Costly
- Time-consuming

Research question:
- Usually active treatment vs. placebo
- Focus on short-term and hard endpoints

Generalisability of results:
- Apply strict exclusion criteria
- Executed under artificial circumstances regarding follow-up and adherence

Statistical considerations:
- Small sample sizes
- Focus on average results

Implications

- Cannot evaluate clinically important long-term endpoints
- Placebo comparison not relevant for clinical practice
- Less focus on soft endpoints, such as patient preferences/ quality of life
- Study conditions do not reflect clinical reality
- Participants are unlikely to be representative for the general T2D patient population
- Lack of data on interactions with concomitant illnesses and treatments
- Subgroup analysis of important patient groups is not possible

Figure 2.1: Relevant limitations and their implications of evidence of RCTs in diabetes research.

Practical issues of RCTs are that they are costly and because of this, their duration is often severely time-limited. It is therefore not possible to study all clinically relevant diabetes end-points which are often rare or end-stage outcomes and develop long-term, such as microvascular outcomes (e.g. retinopathy, nephropathy, or neuropathy). [144, 145] Often, RCTs are also not suitable to address questions and issues clinicians encounter in their practice. In order to find optimal choice for treatment intensification, clinicians are faced with multiple options of active treatments they need to compare. Efficacy and safety trials on the other hand are often designed on the base of placebo comparisons. [148] Additionally, RCTs traditionally focus on hard end points such as physiological measures and disease incidence, but for the treatment of T2D, guidelines emphasise the importance of patient preferences and considerations regarding their quality of life. [147] A major limitation of RCTs to guide treatment decisions is the lack of applicability of trial results on real-life clinical practice. RCTs are conducted under strict study-protocols and artificial circumstances regarding follow-up of participants and adherence incentives, which are implausible in real-life clinical practice. [145, 147] Furthermore, they often apply strong exclusion restrictions regarding factors such as age, comorbidities or T2D duration, and hence, select a homogeneous study population. This population does not reflect the heterogeneous population of T2D patients as seen by clinicians in real-life. [145, 5, 150] For example, a study on the representatives of T2D patient population in large trials of glycaemic control found that trial participants where younger at age of diagnosis compared to the population of individuals with T2D in the UK. [151] Another study found a lack of generalisability for the outcome of cardiovascular benefits of SGLT2i as found in the EMPA-REG OUTCOME trial. [152] When studied in real-world data, the results were only applicable for a small population of individuals with T2D. [153] Lastly, sample sizes of RCTs are usually too low for analysis of important subgroups due to underrepresentation, such as the T2D population of older adults, for example above 65 or 70 years. [154] RCT evidence also focuses mostly on average results, which are not always suitable to help clinicians to make individualized treatment decisions. [145] Young et al. [155] performed a systematic review of meta-analysis studies, RCTs and observational analysis evaluating clinical and biological features associated with treatment effect heterogeneity for SGLT2i and GLP-1. The study concluded that limited evidence is available to reflect on treatment effect heterogeneity in patient subgroups, which is likely due to methodological limitations of published studies.

Due to these shortcomings, it is often not sufficient to make treatment decisions in clinical practice based solely based on RCT evidence. An opportunity for the formulation of treatment guidelines based on a wider spectrum of evidence is to integrate real-world evidence (RWE). RWE is derived for example from longitudinal patient-level data that is not necessarily collected for a specific research purpose. Sources can be electronic healthcare records (EHR) collected by a physician in healthcare setting, insurance claims databases, electronic devices/ software applications or registries. [144, 146] The level of evidence RWE provides depends on many aspects such as data quality, suitability of the data for a specific research question, the rigorous application of the study design, and the choice of appropriate methods for analysis. The latter aspect will invariably refer to the use of a suitable causal inference method as explained in Chapter 1. [144, 156] If all of these aspects are considered appropriately, RWE can reveal insights in comparative effectiveness of treatments under real-world circumstances regarding for example patient behaviour and healthcare practice and the management of a wide range of treatment setting.

With a larger sample size and less restrictive inclusion criteria study populations of real-world data can be more representative, which leads to better generalisability of RWE. Additionally, higher statistical power makes it possible to close knowledge gaps about important subgroups of T2D patients and enables the development of guidelines for individualized treatment of people with T2D. [144] RWE is commonly used in comparative effectiveness and safety studies. Examples of such studies in T2D research are McGovern et al. [157] or Patorno et al. [158] for the study of the safety profile of SGLT2i compared to DPP4i or GLP-1, as relevant comparisons of possible treatment intensification options. Gregg et al. [144] advocate to address the valid concerns of RWE by focusing on reproducibility and transparency and suggest recommendations for standardized reporting of studies based on real-life data. These entail for example the publication of study protocols before analysis, open discussion of methodological criticisms and permitting replication of the research by making the study data available. [144]

2.6 Current evidence on the effectiveness and safety of Sodium-glucose Cotransporter-2 Inhibitors

SGLT2is belong to the novel class of medication for blood glucose management and have been prescribed increasingly over the last years. [126, 127] Additionally, recent changes in treatment guidelines have supported the prescription of SGTL2i for a broad range of people with T2D. [111, 124] These changes were based on evidence of large-scale clinical trials showing that SGLT2is promote cardiorenal protection, reduction of blood pressure, and weight loss independent of glycaemic control. [159, 160, 161] SGTL2is' mode of action operates independently of insulin by inhibiting the reabsorption of glucose in the kidney in case of excess plasma glucose concentration. This leads to increased excretion of glucose in the urine (glycosuria), a mechanism that is associated with weight loss, as well as a reduction in arterial blood pressure due to osmotic diuresis. [162, 163] Furthermore and due to their mode of action, SGLT2is benefit from a reduced hypoglycaemic risk, which is an important advantage for the treatment of older adults with T2D compared to other drug classes with insulin dependent mode of actions. [133, 140] Agents included in the SGLT2i class available in many countries are Canagliflozin, Dapagliflozin, Empagliflozin, and Ertugliflozin. [163]

SGLT2is have well recognized risks that should be considered to optimize the risk-benefit ratio of individualized treatment decisions. [164] Some of the AEs of main concern, especially for the patient population of older adults, are genital infections (GI) and AEs related to diuresis and volume depletion. They are more common in older adults and can lead to confusion and other sequela. [165, 166, 167] Optimal prescribing of SGLT2is in the older patient group requires a full understanding of their risks and benefits but only sparse data is available for patients over 65 years. For example, the average age of two large-scale safety trials on SGLT2i, the EMPA-REG OUTCOME trial and DECLARE-TIMI 58 was 63 (± 9) years and 63.9 (± 6.8) years in the treatment groups respectively. Their results might therefore not be applicable to the real-world population of older adults with T2D. [154] Furthermore, trial studies suffer from small sample sizes for older patients and therefore the results might suffer from potential outlier effects. [165]

Observational studies on this patient population are rare and mostly apply analysis methods which do not account for unmeasured confounding. In the following, evidence on the efficacy and effectiveness as well as the safety of SGLT2is for the general patient population and older T2D patients is summarized.

2.6.1 Important treatment effects

Two main outcomes to study the efficacy and effectiveness of SGLT2i are the reduction of blood glucose levels HbA1c, and weight. Besides blood glucose control, achieving a normal weight in individuals with T2D is one of the most important treatment objectives. Already modest weight loss between 5% and 10% can reduce the risk of diabetes-associated complications and significant weight loss can even potentially resolve the disease. [160, 168] Weight loss has also been investigated as mediator to CVD risk reduction. [160, 169, 170]

Blood glucose control

Evidence of clinical trials for different agents of SGLT2i confirms HbA1c reduction of $4.4 - 12.1$ mmol/mol or $0.4 - 1.1\%$. Furthermore, it was found that the reduction of achieved HbA1c depends on baseline HbA1c, agent and dose. [171, 172, 173, 174] Placebo controlled trial analyses for older T2D patients showed efficacy for SGLT2i in older patients. [175, 176] But another pooled analysis found no difference in efficacy between older and younger patients. [165] More relevant comparisons to guide treatment decisions are between active treatments, as this reflects the necessary choices clinicians have to make in clinical practice. When compared to DPP4i for example, it was found that SGLT2i lead to higher relative effectiveness with a larger reduction in HbA1c. [177] Recent findings of a treatment selection algorithm developed in observational data and validated with trial data also suggest that relative effectiveness of SGLT2i (versus DPP4i) depends on age. [178]

Weight loss

Large-scale clinical trials have shown that SGLT2i reduces body weight. [152, 179] The initial decrease of body weight on SGLT2i can be ascribed to the caloric loss based on glucose excretion and the loss of body water due to osmotic diuresis. [160] The degree of weight loss on SGTL2i has been found to vary across the agents within this treatment class. A meta-analysis of 38 trials concluded for placebo controlled trials average weight loss between 2.66 kg and 1.8 kg depending on the agent. When compared to active comparator treatments (MFN, SU and DPP4i), SGLT2i also showed higher weight loss benefits. [161]

2.6.2 Adverse effects

An increase in urinary glucose excretion due to SGLT2i initiation also leads to a concurrent increase in urinary output, attention must therefore be paid to the occurrence of osmotic symptoms such as dehydration and genital infections. [180] Other potential AEs might be rare but life-threatening such as DKA. Lastly, falls and amputations are considered in this summary. Not all trial evidence identified these AEs for SGLT2is, but further insights about possible connection with SGLT2i initiation is crucial, as they can have severe consequences on the quality of life of T2D patients, and might be more common in older adults. Recent evidence suggest increased risk for short-term treatment discontinuation of SGLT2i in older adults, which is a proxy for tolerability of the treatment. [178]

Osmotic diuresis-related adverse effects

Genital infections are well-recognized AEs of SGLT2is due to induced glycosuria. [164] They were found to be the most common AEs in clinical trials, but were generally mild to moderate and did not lead to discontinuation of SGLT2i. [152, 164, 179, 181, 182] These results have been confirmed by observational studies including a broader range of T2D patients. Furthermore, female gender and a history of genital infection previous to SGTL2i initiation have been found to be risk factors for this adverse effect. [157, 183] The observational study by Goldman et al. [154] on the safety profile of SGLT2i in T2D patients 75 years or older utilized safety reports from the US Food and Drug Administration (FDA) global database and reported sex-adjusted odds ratios. The study did not find risk differences between the age groups regarding genital infections. But results from this study are not given as true incidences because the number of patients exposed to the drugs are not known. The incidence rate of adverse effects in the older patient population could therefore still be higher. [154]

Other osmotic diuresis-related AEs are volume depletion, micturition control and urinary frequency. Volume depletion can lead to dehydration due to osmotic diuresis. Furthermore, it can lead to postural dizziness and orthostatic hypertension which can be of special concern for older patients with kidney disease or those on loop diuretics. [162, 164] Trial evidence confirmed that volume depletion are more common on SGTL2i (canagliflozin and empagliflozin) [152, 179] and further study results with placebo controlled trial data indicate an increase in events for older patients [165, 184, 185]. The retrospective, pharmacovigilance study of the FDA's global database of safety reports showed that dehydration was significantly more common in patients initiating SGLT2i but no difference between the age groups of younger and older patients (≥ 75 years) was detected. [154] Little studies are available on urinary frequency, but patients may present with frequent urination or overactive bladder syndrome following the initiation of SGLT2i. Especially, daytime frequency has been shown to be increased in patients without previous overactive bladder syndrome. In patients with established overactive bladder previous to SGLT2i initiation quality of life and overactive bladder was not negatively affected according to a before–after comparative study. [180]

Diabetic ketoacidosis

DKA is a rare but serious complication of diabetes and can be life-threatening. It is characterized by ketonaemia, acidaemia and hyperglycaemia, even though raised blood glucose levels do not always need to be present. [186, 187] The mechanism with which SGLT2i might cause DKA is not fully elucidated but might be a consequence of SGLT2is noninsulin-dependent glucose clearance, hyperglucagonemia, and volume depletion. [164, 188] Older patients may have long-duration diabetes and thereby a low residual insulin secretion, a condition that is known to increase the risk of DKA. Furthermore, individuals from this patient group may present more frequent acute complications such as infections which are also recognised to increase the risk of DKA. [162] Reports of a possible association between SGLT2i and DKA has prompted the FDA [189] and the European Medicines Agency (EMA) [190] to issue warnings for this treatment class. The suspected association between SGLT2i initiation and DKA has not been confirmed in large-scale clinical safety trials such as the EMPA-REG OUTCOME trials [152] or CANVAS [179]. Real-life evidence for example from a large claims database of commercially insured patients in the US showed significant association in a relative risk study comparing SGLT2i with DPP4i and the application of PSM. Adjusted hazard ratios (HR) for the treatment group with 180 days follow-up was 2.2 [CI $95\% \, 1.4, 3.6$]. But average age in this study was low with 54.8 ± 9.4 for individuals on SGLT2i and 54.4 ± 10.8 for individuals on DPP4i. [191] Another register based study comparing SGLT2i and GLP-1 also showed a significantly increased relative risk for DKA, even though it remained a rare outcome. [192] The observational study by Goldman et al. [154] focused on a comparison of older and younger patients and did find an increased risk in both groups but no group differences. Further studies on DKA risk of SGLT2is are needed to fully judge the possibility of an association. SGLT2i should be avoided for patients with risk factors for ketoacidosis and needs to be stopped in case of a DKA event. Furthermore, SGLT2i should not be started again immediately after a DKA event as reports indicate reoccurring DKA events after continuation of SGLT2i. [162, 166]

Further adverse effects

Falls are generally a concern for diabetic patients as they can lead to disabilities and a lower quality of life. Additionally, older individuals with T2D are of increased risk of falling for example due to greater impairments in posture and gait. [193, 194] Events of falls, especially for older adults, may also be caused due to dehydration and dizziness, when initiating SGLT2i. [154] Studies including falls as AE outcome are rare, but a disproportionality analysis of SGLT2i compared to other non-insulin antidiabetic drugs from FDA safety reports found slightly elevated sex-adjusted reporting odds ratio. Results of this study showed no difference between the younger and older patient group (≥ 75 years). [154]

Another potential AE of SGLT2i are amputations. Lower extremity amputations (leg, foot, and toe) have been reported most often in connection with SGLT2i. The risk of amputation for patient initiating SGLT2i is still unclear and study results are heterogeneous for different agents in this treatment class. [162, 164, 195] In the CANVAS trial a significant association of Canagliflozin with amputations was found, with the highest absolute risk in patient with history of amputation and peripheral vascular disease. [179] An association was not found for Empagliflozin in the EMPA-REG OUTCOME trial. [152] Based on reports of the CANVAS trial, the FDA [196] issues a warning for Canagliflozin and the EMA [197] issues a warning for all SGLT2i agents. Some observational studies also found increased risk of amputations when comparing SGLT2i with DPP4i [198] and one study comparing SGLT2i with GLP-1 [199] found that amputations were more apparent in the group of patients older than 65 years and in patients with baseline CVD. In clinical practice SGLT2i should be prescribed with caution to patients with risk factors of amputations, for example older patients or patients with vascular disease, diabetic foot ulcer, or previous amputations. [162]

2.7 Profile of the Clinical Practice Research Datalink

The RWE dataset used for the applied research in this thesis comes from Clinical Practice Research Datalink (CPRD), a large longitudinal patient level database of anonymised EHR which have been collected in routine primary care practice by general practitioners in the UK. Data has been collected since 1987 from patients registered at general practices (GPs) on aspects relevant for health care research such as diagnoses, symptoms, prescriptions, referrals, tests, demographics, and behavioural factors. [200] This large high-quality research database is well suited for effectiveness and safety studies of T2D treatments as it provides representative population data with comprehensive capture of risk factors and outcomes. [201] CPRD includes 11.3 million patients from 674 GPs, 4.4 million of which are currently alive. This means that around 6.9% of UK population are covered by CPRD. The patient population has been found to be representative of the UK general population regarding age, sex and ethnicity. [200]

Another benefit besides its size and representativeness of CPRD is the enrichment of the database through data linkages from secondary care, disease specific cohorts and mortality records. Data linkage is available for 75% of English GPs, which make up 58% of GPs participating in CPRD. Linkage to other data sources is done by the Health and Social Care Information Center as third party. Linked information includes hospital data from the Hospital Episode Statistics (HES), mortality data including causes of death from the Office for National Statistics (ONS), deprivation data such as the Index of Multiple Deprivation and Townsend score, and disease registries, for example the National Cancer Intelligence Network and the Myocardial Ischaemia National Audit project. [200, 202] This enrichment of the primary care EHR will provide a more all-encompassing picture of individual patient care pathways. Furthermore, CPRD data is subjected to over 900 quality checks which cover integrity, structure, and format of the data. [202]

Aspects that are important to consider when using CPRD data for health research is missing data in the primary care records which are manually recorded by general practitioners. Furthermore, some conditions which can be treated with over-the-counter medication or home remedies will not be recorded in CPRD, but might be important in order to correctly identify outcomes of interests. Human errors in the process of data entry and misclassifications are also possible. Lastly, GP IT systems might vary and coding between GPs and over time might therefore not be the same. [202] Furthermore, as prescriptions in CPRD data are not randomized, it is essential to employ suitable causal inference methods as outlined in Chapter 1 to minimize the risk of confounding bias in the treatment effect estimation.

Two CPRD datasets, CPRD Gold (download July 2019) and CPRD Aurum (download November 2020) will be used in this thesis to study the effectiveness and safety profile of SGLT2i. The CPRD Aurum dataset also includes data linkages to the described data sources. The T2D cohorts were built using a published protocol. [203] Individuals with T2D were identified using the presence of a diagnostic code for diabetes and the prescription of one or more glucose lowering medications. T1D and other types of diabetes where excluded. Individuals with possible T1D were identified if their age was less than 35 at diagnosis, their only treatment was insulin or insulin was initiated 1 year after diagnosis. Other diabetes forms such as gestational diabetes and monogenic diabetes were identified with indicative diagnostic codes. For this identification process, the date of diagnosis was defined by the earliest date of a diabetes diagnostic code, indication of glucose lowering medication, or glycated hemoglobin (HbA1c) $\geq 6.5\%$ (48 mmol/mol). Individuals were also excluded if their first diabetes indicator date fell within 3 months of practice registration, as their date of diagnosis was uncertain. [157, 203] Diagnosis of comorbidities, adverse effects and treatment prescriptions were identified using diagnosis codes. Lists for these codes were developed by a clinician with expertise in clinical code list development for research. The utilized code lists are published at: `https://github.com/Exeter-Diabetes/CPRD-Codelists`. Further description of the datasets will be given in more detail in the respective chapters.

2.8 Conclusion

The management of T2D is complex and requires clinicians to take many factors about individual circumstances and preference of patients into account. Treatment guidelines are based mainly on RCT evidence about the efficacy and safety profile of treatment choices. SGLT2i is an important treatment class due to its weight loss benefit and cardiorenal protective effects. Prescribing of SGLT2i has increased in recent years and after recent treatment guideline changes a large proportion of T2D patients have become eligible for this treatment class. More information about the safety profile for SGLT2i are needed especially for the patient population of older adults, as possible AEs associated with SGLT2i could have severe consequences for this patient group. RWE needs to be integrated into the decision making process of clinicians, as RCTs have limitations regarding for example the under-representation of patient subgroups such as older adults, and are not able to answer all questions clinicians face in real-life practice in particular regarding the risk of less common AEs. CPRD provides a rich data source to gain RWE on the effectiveness and safety for comparisons of T2D oral agents relevant in clinical practice. But in order to derive high quality RWE from this observational data, adequate causal inference methods need to be employed.

Chapter 3

Triangulating Instrumental Variable, confounder adjustment and difference-in-difference methods for comparative effectiveness research in observational data

Laura M. Güdemann, John M. Dennis, Andrew P. McGovern, Lauren R. Rodgers, Beverley M. Shields, William Henley & Jack Bowden
on behalf of the MASTERMIND consortium

Planned submission in: BMC Medical Research Methodology,
for the special issue on causal inference and observational data

Pre-published at: `https://arxiv.org/abs/2202.09164`

Author contribution

LMG and JB conceived and developed the methodological framework, with the constant supervisory support of BMS, LRR and JMD. APM provided invaluable clinical insight for the applied analyses in Section 3.6 and WH provided useful insight on the assumptions required by DiD regression. LMG conducted the analysis and drafted the original version of the paper which all authors helped to edit. All authors read and approved the final version of the manuscript.

3.1 Abstract

Observational studies can play a useful role in assessing the comparative effectiveness of competing treatments. In a clinical trial the randomization of participants to treatment and control groups generally results in well-balanced groups with respect to possible confounders, which makes the analysis straightforward. However, when analysing observational data, the potential for unmeasured confounding makes comparing treatment effects much more challenging. Causal inference methods such as Instrumental Variable and prior event rate ratio approaches make it possible to circumvent the need to adjust for confounding factors that have not been measured in the data or measured with error. Direct confounder adjustment via multivariable regression and propensity score matching also have considerable utility. Each method relies on a different set of assumptions and leverages different aspects of the data.

In this paper, we describe the assumptions of each method and assess the impact of violating these assumptions in a simulation study. We propose the prior outcome augmented Instrumental Variable method that leverages data from before and after treatment initiation, and is robust to the violation of key assumptions. Finally, we propose the use of a heterogeneity statistic to decide if two or more estimates are statistically similar, taking into account their correlation. We illustrate our causal framework to assess the risk of genital infection in patients prescribed Sodium-glucose Cotransporter-2 Inhibitors versus Dipeptidyl peptidase-4 Inhibitors as second-line treatment for type 2 diabets using observational data from the Clinical Practice Research Datalink.

Keywords: causal inference, unmeasured confounding, triangulation, Instrumental Variable method, prior event rate ratio approach

3.2 Introduction

The gold standard approach for evaluating the efficacy of treatments is a randomized controlled trial (RCT). Due to strict specifications of RCTs with regard to blinding and randomization of treatment assignment, causal conclusions about the treatment's effect on patient outcomes can be drawn without the need to adjust for prognostic factors, since they should be well-balanced across trial arms. This remains true even if the trial is affected by non-adherence, and non-adherence is predicted by the aforementioned prognostic variables. [6]

Observational data, for example from electronic healthcare records, provide vital means for assessing the comparative effectiveness of commonly prescribed medications with similar indications. Since these data are collected as part of routine care, treatment assignment is not randomized. This opens up the possibility that treatment choice (or the extent of treatment received) and patient outcomes may be simultaneously predicted by common variables, which could bias the analysis because of a lack of balance across treatment groups, in contrast to the adherence affected RCT setting previously discussed. This phenomenon is referred to as 'confounding' and we will refer to such common factors as confounders from now on. [6, 5, 204]

Standard causal inference methods such as stratification, multivariable regression or propensity score matching make it possible to analyse observational data and draw causal conclusions as long as all confounders can be accurately measured and appropriately adjusted for. [6] For example, Dawwas et al. [205] used propensity score matched data for a retrospective cohort study for a comparative risk analysis of cardiovascular outcomes in people with type 2 diabetes (T2D) initiating Dipeptidyl peptidase-4 inhibitors (DPP4i) versus Sodium-glucose Cotransporter-2 Inhibitors (SGLT2i) therapy. Another example using standard causal inference methods is McGovern et al. [157] who used multivariable Cox regression and propensity scores to define important clinical groups of people with T2D initiating either DPP4i or SGLT2i, with high risk of genital infection.

Failure to measure and appropriately adjust for all confounders can bias the estimation of the true causal effect of treatment on the outcome of interest. Two causal inference approaches which circumvent the problem of unmeasured confounding are the Instrumental Variable (IV) and the prior event rate ratio (PERR) method. The IV approach addresses confounding by substituting each patients' observed treatment with a predicted treatment. This prediction is made using a variable (the instrument) that is assumed to be independent of any confounders and only affects the outcome through the treatment. Randomization to a treatment group within a RCT is perhaps the best example of an IV, and can therefore be used to adjust for non-adherence. [6, 55, 206] Because of this, IV analyses using observational data are generally equated with the creation of a pseudo-randomized controlled trial. Examples of IVs for observational data include geographic information such as the distance to the nearest health facility [207], germ line genetic information [208] or healthcare providers' preference for a particular treatment [92]. In this paper we will subsequently construct an IV of this latter type.

The prior event rate ratio method is an alternative quasi-experimental approach which leverages data at two time points. Specifically, the outcome must be measured in the 'prior' period before initiation of treatment and then in the 'study' period after treatment has commenced. [50] The treatment effect is first estimated in the prior period by (somewhat paradoxically) regressing the prior outcome on the study period treatment indicator. This is assumed to capture the degree of unmeasured confounding in the treatment effect subsequently estimated in the study period, which can then be subtracted out to de-bias the analysis. The approach relies on the assumption of time invariant unmeasured confounding across both time periods. For related reasons it is necessary that a patient's outcome in the prior period does not influence the allocation of the study period treatment. Furthermore, the prior and study event of interest should be of the same nature and non-terminal, such as death. [49, 50]. The PERR method is generally applied to time-to-event data, but is directly analogous to the method of difference-in-difference (DiD) regression in the case of continuous or binary outcomes.

Figure 3.1 shows a causal diagram illustrating the possible relationship between: the outcome in the prior and study periods (Y_0 and Y_1 respectively); the treatment indicator (X); an IV (Z); measured confounders of treatment and outcome in both periods ($\mathbf{W_0}$ and $\mathbf{W_1}$ respectively); and unmeasured confounders ($\mathbf{U}$). The measured and unmeasured confounder are summarized in matrices of sizes $\mathsf{N} \times \mathsf{G}$ and $\mathsf{N} \times \mathsf{M}$ respectively. In this diagram, the assumptions related to variable relationships of both the DiD (analogous to PERR) and IV approaches are satisfied, but the 'no unmeasured confounder' assumption underlying a direct adjustment strategy is not.

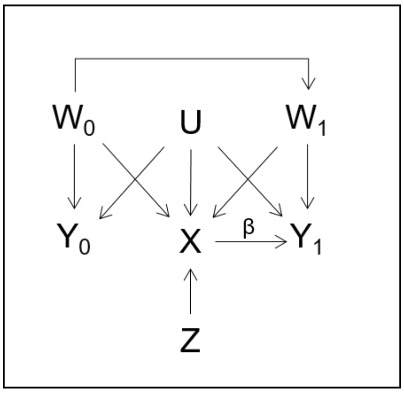


Figure 3.1: Causal diagram showing the relationship between Y_0*,* Y_1*,* X*,* $\mathbf{U}$*,* $\mathbf{W_0}$ *and* $\mathbf{W_1}$ *in the case where both the IV and DiD assumptions are satisfied. The estimates and assumptions are explained in detail in Section 3.3.*

In this paper we consider the joint application of direct confounder adjustment, IV and DiD approaches for estimating the causal effect of treatment using observational data. In Section 3.3 we give a more detailed description of each method and introduce a heterogeneity statistic to decide if two or more estimators are sufficiently similar. In Section 3.4 we assess the performance of these approaches in a detailed simulation study. In Section 3.5 we consider an extension of the standard IV approach using pre- and post-treatment outcome data that can be used in scenarios where the assumptions of both the standard IV and DiD approaches are violated. The method utilizes an interaction term of the instrument (Z) and the outcome variable measured before treatment initiation (Y_0), as new Instrumental Variable. Hence, the method is able under to allow for an influence of Y_0 on X and Z on Y_1. We call this extension the prior outcome augmented Instrumental Variable method (POA-IV). In Section 3.6 we apply our methods to routinely collected healthcare records to assess the causal effect of SGLT2i compared to DPP4i as second-line therapy on the risk of genital infections, exploiting variation in healthcare provider prescribing habits to construct an IV. We conclude in Section 3.7 with a discussion and point to further research. Source code for this research for all simulations and the application case study in this paper is available at `https://github.com/GuedemannLaura/POA-IV`.

3.3 Methods

We are interested in estimating the comparative effectiveness of two treatments ($X = 1$ compared to $X = 0$) on outcome Y_1 using observational data. The target of this analysis is a hypothetical estimand:

$$\beta = E[Y_1|X = 1] - E[Y_1|X = 0] \tag{3.1}$$

That is, the difference in expected outcomes if all patients could receive treatment level 1 compared to treatment level 0. For simplicity, we will assume in this section that the outcome of interest is continuous, with the extension to binary outcomes discussed in Section 3.3.4.

3.3.1 The 'as Treated' and 'Corrected as Treated' estimate

In a RCT with complete adherence to the assigned treatment, hypothetical estimand β could be consistently estimated using the 'as Treated' estimate, by comparing the average outcome across both treatment groups. Using the potential outcome notation, let $Y_{1\,i}^{X_i=x}$ denote the outcome of patient i if assigned treatment level $X_i = x$ and Tx, Ct referring to treatment group and control group respectively. The as Treated estimate can be written as:

$$\hat{\beta}_{aT} = \bar{Y}_{1\,i\in Tx}^{X=1} - \bar{Y}_{1\,i\in Ct}^{X=0},$$

with $\bar{Y}$ denoting the sample average. Complete adherence could be illustrated in Figure 3.1 by letting Z represent the randomized treatment assignment and removing all arrows into X from W_0, W_1 and U, so that only the $Z \to X$ arrow remains. Difficulties emerge when calculating the as Treated estimate with observational data, because treatment assignment is not randomized or controlled by the researcher. It is then possible that factors exist which simultaneously affect (or confound) the treatment assignment and the outcome. This would lead to an imbalance across the treatment groups with respect to W_0, W_1 and U and the estimate $\hat{\beta}_{aT}$ will consequently be biased due to confounding.

If all confounding factors are known and can be appropriately measured and adjusted for - which we call the ‘no unmeasured confounder’ (NUC) assumption - a ‘Corrected as Treated’ (CaT) estimate that is additionally adjusted for these factors can consistently estimate β. Returning to Figure 3.1, if the NUC assumption held so that $\mathbf{U}$ was absent from the diagram it would be sufficient to adjust for $\mathbf{W_1}$ since $\mathbf{W_0}$ only affects $\mathsf{Y_1}$ through $\mathbf{W_1}$ and the CaT estimate would be

$$\hat{\beta}_{\mathsf{CaT}} = \hat{\mathsf{E}}[\mathsf{Y}_1|\mathsf{X}=1,\mathbf{W_1}] - \hat{\mathsf{E}}[\mathsf{Y}_1|\mathsf{X}=0,\mathbf{W_1}]. \tag{3.2}$$

This could be estimated from fitting the following multivariable regression of Y_1 on X and $\mathbf{W_1}$ as

$$\mathsf{E}[\mathsf{Y}_1|\mathsf{X},\mathsf{W}_1] = \beta_{\mathsf{Y}_1,0} + \beta_{\mathsf{CaT}}\mathsf{X} + \beta_{\mathbf{Y_1},\mathbf{W_1}}\mathbf{W_1},$$

where $\beta_{\mathbf{Y_1},\mathbf{W_1}}$ is a $\mathsf{G} \times 1$ column vector and $\mathbf{W_1}$ a matrix of size $\mathsf{N} \times \mathsf{G}$.

3.3.2 The Instrumental Variable estimate

In many settings the NUC assumption may be thought unreasonably strong. The Instrumental Variable (IV) method offers a means for circumventing the problem of unmeasured confounding to consistently estimate the hypothetical estimand. It works via the construction of a pseudo-randomized controlled trial using a variable Z which needs to fulfil the following three assumptions in order to be a valid IV:

- IV1: Z is associated, or predicts X;
- IV2: Z is independent of Y_1 given $\mathbf{W_0}$, $\mathbf{W_1}$, X and $\mathbf{U}$;
- IV3: Z and Y_1 do not share a common cause.

IV1 is often referred to as the relevance assumption and the $\mathsf{Z}-\mathsf{X}$ relationship can be empirically tested from a regression of X on Z. The assumption would be invalidated if this association is weak, with an F-statistic of at least 10 often used as a threshold for good strength of an IV. [77] Assumption IV2 is also referred to as the exclusion restriction and requires Z to only influence Y_1 through X but not directly. IV3, the exchangeability assumption, requires that Z and Y_1 are not themselves confounded. [68, 70]

As explained in Chapter 1.8.4, the IV estimate for β can be estimated as the ratio of the $Y_1 - Z$ association and the $X - Z$ association for a binary IV and continuous outcome with no adjustments for covariates made:

$$\hat{\beta}_{IV} = \frac{\hat{E}[Y_1|Z = 1] - \hat{E}[Y_1|Z = 0]}{\hat{E}[X|Z = 1] - \hat{E}[X|Z = 0]} \quad . \tag{3.3}$$

In order to enable consistent estimation of hypothetical estimand (3.1) using (3.3), we additionally make the homogeneity assumption that the average treatment effect is constant across both levels of the IV Z, at each level of the treatment [206]:

$$E[Y_i(X = 1) - Y_i(X = 0)|Z = 1, X = x] = E[Y(X = 1) - Y(X = 0)|Z = 0, X = x].$$

A more general method for IV estimation with a continuous outcome that allows for multiple IVs and covariate adjustment is Two-Stage Least Squares (TSLS) [68]. To implement TSLS with a single IV Z and the measured confounder $\mathbf{W_1}$, we first fit a logistic regression model for X given Z and $\mathbf{W_1}$:

$$\text{Logit}(\Pr[X = 1|Z, \mathbf{W_1}]) = \beta_{X,0} + \beta_{X,Z}Z + \beta_{\mathbf{X},\mathbf{W_1}}\mathbf{W_1} \tag{3.4}$$

where $\beta_{\mathbf{X},\mathbf{W_1}}$ is a $G \times 1$ column vector and $\mathbf{W_1}$ a matrix of size $N \times G$. The estimated coefficients of this model are then used to predict X given Z and $\mathbf{W_1}$ as $\hat{X}$, which is then itself regressed on Y_1 and $\mathbf{W_1}$ in a second-stage model:

$$E(Y_1|\hat{X}, \mathbf{W_1}) = \beta_{Y_1,0} + \beta_{IV}\hat{X} + \beta_{\mathbf{Y_1},\mathbf{W_1}}\mathbf{W_1} \tag{3.5}$$

The coefficient of $\hat{X}$ is then taken as the TSLS estimate [70]. Using a valid IV, the TSLS is consistent under the homogeneity assumption and additionally that the covariates are correctly modelled in (3.5) [71, 72].

3.3.3 Difference-in-difference estimate

An alternative approach to adjust for unmeasured confounding is the difference-in-difference (DiD) estimate. It can be applied to continuous and binary outcomes and is conceptually equivalent to the prior event rate ratio (PERR) method typically applied to time-to-event outcomes [55]. Borrowing the terminology of the

PERR approach, DiD estimation leverages data from two periods: the prior period before drug initiation and the study period after drug initiation. For the estimation of the treatment effect in the study period, the treatment effect measured for the prior period is used to capture the degree of unmeasured confounding. The method presumes that the treatment effect measured in the prior period reflects the composite effect of measured and unmeasured confounders on the outcome, if none of the participants receive any of the study treatments in the prior period. [48, 49, 50, 209] Once estimated, it can then be subtracted from the $X - Y_1$ association (or the as Treated estimate). This approach relies on the following assumptions:

- DiD1: Y_0 does not influence the treatment decision X directly
- DiD2: the effect of U on the outcome is constant across time conditional on $\mathbf{W_0}$ and $\mathbf{W_1}$. [50, 49]

Previous studies show that the DiD method is biased in case of the violation of assumption DiD1 [54, 210] and DiD2 [52]. A formal proof that these assumptions are sufficient for identification of hypothetical estimand (3.1) is given in Appendix 3.2. In Figure 3.1 assumption DiD1 is satisfied and throughout all presented studies we assume that DiD2 is satisfied.

For continuous outcomes the DiD estimate can be calculated by subtracting the results of two linear regressions from the prior and study period:

$$\begin{aligned}\hat{\beta}_{\text{DiD}} = \hat{\text{E}}[Y_1|X = 1, \mathbf{W_1}] - \hat{\text{E}}[Y_1|X = 0, \mathbf{W_1}] \\ -(\hat{\text{E}}[Y_0|X = 1, \mathbf{W_0}] - \hat{\text{E}}[Y_0|X = 0, \mathbf{W_0}]).\end{aligned} \tag{3.6}$$

The DiD estimate can also be calculated for a sample of N individuals via the following single regression model:

$$\begin{aligned}\text{E}[Y^*|X^*, W^*, P^*] = \beta_{Y,0} + \beta_{Y^*,P^*}P^* + \beta_{Y^*,X^*}X^* + \beta_{\text{DiD}}P^* \cdot X^* + \\ \beta_{Y^*,W^*}\mathbf{W}^* + \beta_{Y^*,W^*P^*}\mathbf{W}^* \cdot P^*.\end{aligned} \tag{3.7}$$

Here, $X^* \in \{0, 1\}$ and $P^* \in \{0, 1\}$ are 2N-length treatment and period indicator variables and the variables $Y^* = (Y_0, Y_1)^T$, $\mathbf{W}^* = (\mathbf{W_0}, \mathbf{W_1})$ summarize the information of outcomes and covariates for both periods in a vector of the same size.

The regression coefficients of the $P^* \cdot X^*$ interaction term is taken as the DiD estimate. [57] Fitting this model facilitates the easy extraction of a standard error for the DiD estimate directly from the hessian matrix.

The DiD method utilizes only two outcome measurements before and after treatment initiation and can be viewed as a simple special case of an interrupted time series analysis, which incorporates data from multiple time points within a formal longitudinal model [211, 212, 213, 214]. Due to the limited availability of repeated outcome and confounder measurements to two periods (before and after treatment initiation) in our data and our focus on triangulating findings across methods, we restrict our attention to the DiD approach in this paper.

3.3.4 Extension to binary outcomes

In case of a binary outcome the comparative treatment effect can be estimated using the CaT, IV, CF and DiD models using logit or probit models, instead of linear regressions for continuous outcomes. Whilst TSLS is the standard tool for IV analysis with continuous outcomes, the control function (CF) method is typically used for binary outcomes. This method can accommodate linear and non-linear associations between the IV and treatment and between the treatment and the outcome [68, 75]. For the first stage model and continuous outcomes the IV is regressed on treatment and all measured confounders, as shown in model (3.4). From this regression the residual $\hat{\Delta} = X - \hat{X}$ is calculated and used in the second stage model with

$$\begin{aligned}\text{Logit}(\Pr[Y_1 = 1|X, Z, \mathbf{W_1}, \hat{\Delta}] = \beta_{Y_1,0} + \beta_{Y_1,W_1}\mathbf{W_1} + \beta_{CF}X \\ + (\beta_{Y_1,\hat{\Delta}} + \beta_{Y_1,Z\hat{\Delta}}Z)\hat{\Delta}.\end{aligned} \tag{3.8}$$

We will use both the standard IV and CF approaches for IV analyses in this chapter.

When employing logistic regression models for the CaT, IV, CF and DiD models, for our purposes we prefer to extract the treatment estimate as a risk difference, or an Average Marginal Effect (AME) [206]. For example, in the case of the CF estimate, after fitting model (3.8) to obtain estimates for its constituent parameters, the AME is calculated as the difference in average predicted probabilities when X is fixed at 1 and 0 respectively:

$$\hat{\beta}_{\mathsf{CF}} = \frac{1}{N}\sum_{i=1}^{N}\left\{\widehat{\Pr}[Y_1 = 1|X = 1, Z, \mathbf{W_1}, \hat{\Delta}] - \widehat{\Pr}[Y_1 = 1|X = 0, Z, \mathbf{W_1}, \hat{\Delta}]\right\}$$

In R, this can easily be done using the `margins()` package [215]. Choosing a scale that is collapsible makes it more straightforward to compare estimates across different methods which use different covariate adjustment sets. Non-collapsible measures of associations differ in magnitude over levels of measured confounders when adjusted versus not adjusted for measured confounders. [6] Even if their respective assumptions are all satisfied, using a non-collapsible scale such as an odds ratio could mean that the underlying causal estimands of two methods are in fact distinct. [216]

3.3.5 Similarity statistic

In order to assess the similarity of the estimates, after taking care to estimate them on the same scale and whilst accounting for their correlation, we use a generalized heterogeneity statistic [217] of Cochran's Q [218] of the form

$$Q_e = (\hat{\boldsymbol{\beta}}_e - \hat{\beta}_{\mathsf{IVW},e})\hat{\Sigma}_e^{-1}(\hat{\boldsymbol{\beta}}_e - \hat{\beta}_{\mathsf{IVW},e})^T \tag{3.9}$$

where

- e is the set of estimates, for example $\{\mathsf{CaT}, \mathsf{IV}, \mathsf{DiD}\}$;
- $\hat{\boldsymbol{\beta}}_e$ is a vector of all estimates in e with s-th entry $\hat{\beta}_{es}$;
- $\hat{\beta}_{\mathsf{IVW},e}$ is the inverse variance weighted average of all estimates in e;
- $\hat{\Sigma}_e$ is the covariance matrix for $\hat{\boldsymbol{\beta}}_e$, approximated by a non-parametric bootstrap;
- $\hat{\beta}_{\mathsf{IVW},e}$ is calculated using w_{es} the inverse variance of the corresponding estimate and

$$\hat{\beta}_e = \frac{\sum_{s\in e} w_{es}\hat{\beta}_{es}}{\sum_{s\in e} w_{es}}.$$

Under the assumption that all estimates in e are targeting the same underlying quantity, Q_e is asymptotically $\chi^2_{n_e-1}$ distributed, n_e indicating the number of estimates in the set e. This assumption is rejected at level α if $Q_e > \chi^2_{n_e-1}(1-\alpha)$

and the estimates are assumed not to be similar. Equation (3.9) can therefore be used to assess the extent of agreement across estimators and, by extension, the validity of the assumptions that they rest on. This approach can be seen as a generalisation of the causal triangulation framework for uncorrelated estimates described in Bowden et. al. [11]. We showcase an application of the Q_e in Section 3.4.3.

3.4 Simulation Study

In this section we employ a Monte-Carlo simulation to study the performance of the CaT, IV (TSLS and CF) and DiD estimates in scenarios where their specific assumptions are variously satisfied and violated. The simulation was conducted in R Studio (version 4.1.2) and the set-up is motivated to a degree by the applied analysis in Section 3.6.

3.4.1 Simulation set up

Across 1000 independent simulations, data is generated for $N = 5000$ patients grouped into $J = 50$ clusters, with each cluster representing a healthcare provider (e.g. a general practitioner). The full data generating models are summarized in Appendix 3.1 but the main features are now described. The treatment group indicator X and the outcome variables, Y_0 and Y_1 are simulated as binary variables, in each case representing the presence or absence of a binary adverse event. The true treatment effect, quantified on the risk difference scale, $\beta = 0.1$. Therefore, the average causal effect of treatment 1 versus 0 is a 10% increase in adverse event risk. Further information on the chosen parameter values can be found in the additional provided material at `https://github.com/GuedemannLaura/POA-IV`.

Treatment and outcome variables are allowed to depend in principle on: measured confounders, $\mathbf{W_0}$ and $\mathbf{W_1}$; one unmeasured confounder $\mathbf{U}$ (normally distributed); and the IV Z (simulated as a binary variable). Specifically, Z is constant at the healthcare provider level, and therefore conveys information about providers' preference to prescribe one treatment over the other. Here it is assumed that provider preference is known and no proxy variable construction is needed.

Figure 3.2 summarizes the 8 scenarios implemented in the simulation. In scenario 1, the NUC, IV and DiD assumptions are all satisfied. In scenarios 2-4, the NUC assumption is satisfied, but certain IV and DiD assumptions are violated. Specifically, in Scenario 2, the DiD1 assumption which requires X to be independent of Y_0 is violated. This could be, for example, because the occurrence of an adverse event in the prior period influences the providers' treatment decision in the study period. In scenario 3, IV2 (exclusion restriction) is violated. This could for example represent the case where providers' preference for treatment is associated with their tendency to record adverse events. Previous knowledge about adverse events for a specific treatment could easily give rise to this effect. Both the IV2 and the DiD1 assumptions are violated simultaneously in scenario 4. In scenarios 5 to 8 unmeasured confounding is present (NUC violated) although it is constant over the two time periods. In addition to the unmeasured confounding, the DiD1 and IV2 assumptions are violated in scenarios 6 and 7 respectively. In Scenario 8 the NUC, IV2 and DiD1 assumptions are all violated. The eight scenarios are illustrated using causal diagrams in Figure 3.2. The CaT, IV, CF and DiD estimates were calculated by fitting the models listed in Table 3.1 using logistic regression in tandem with the `margins()` package [215], as described in the previous section. As IV2 is violated in scenarios 3, 4, 7 and 8, Z becomes a measured confounder of Y_1. Is it therefore included in the CaT and DiD model for these scenarios.

Estimate		Fit
CaT	$\hat{\beta}_{\mathsf{CaT}}$	$\mathsf{Y}_1 \sim \mathsf{X} + \mathsf{W}_1$
IV	$\hat{\beta}_{\mathsf{IV}}$	First stage model: $\mathsf{X} \sim \mathsf{W}_1 + \mathsf{Z}$ Second stage model: $\mathsf{Y}_1 \sim \hat{\mathsf{X}} + \mathsf{W}_1$
CF	$\hat{\beta}_{\mathsf{CF}}$	First stage model: $\mathsf{X} \sim \mathsf{W}_1 + \mathsf{Z}$ Second stage model: $\mathsf{Y}_1 \sim \mathsf{X} + \mathsf{W}_1 + \hat{\Delta} + \hat{\Delta} \cdot \mathsf{Z}$
DiD	$\hat{\beta}_{\mathsf{DiD}}$	$\mathsf{Y}^* \sim \mathsf{P}^* + \mathsf{X}^* +^* \mathsf{P} \cdot \mathsf{X}^* + \mathsf{W}^* + \mathsf{W}^* \cdot \mathsf{P}^*$

Table 3.1: Summary of the models for CaT, IV, CF and DiD fitted in the simulation. For scenarios 3, 4, 7, 8 Z *is included in the DiD and CaT model as measured confounder.*

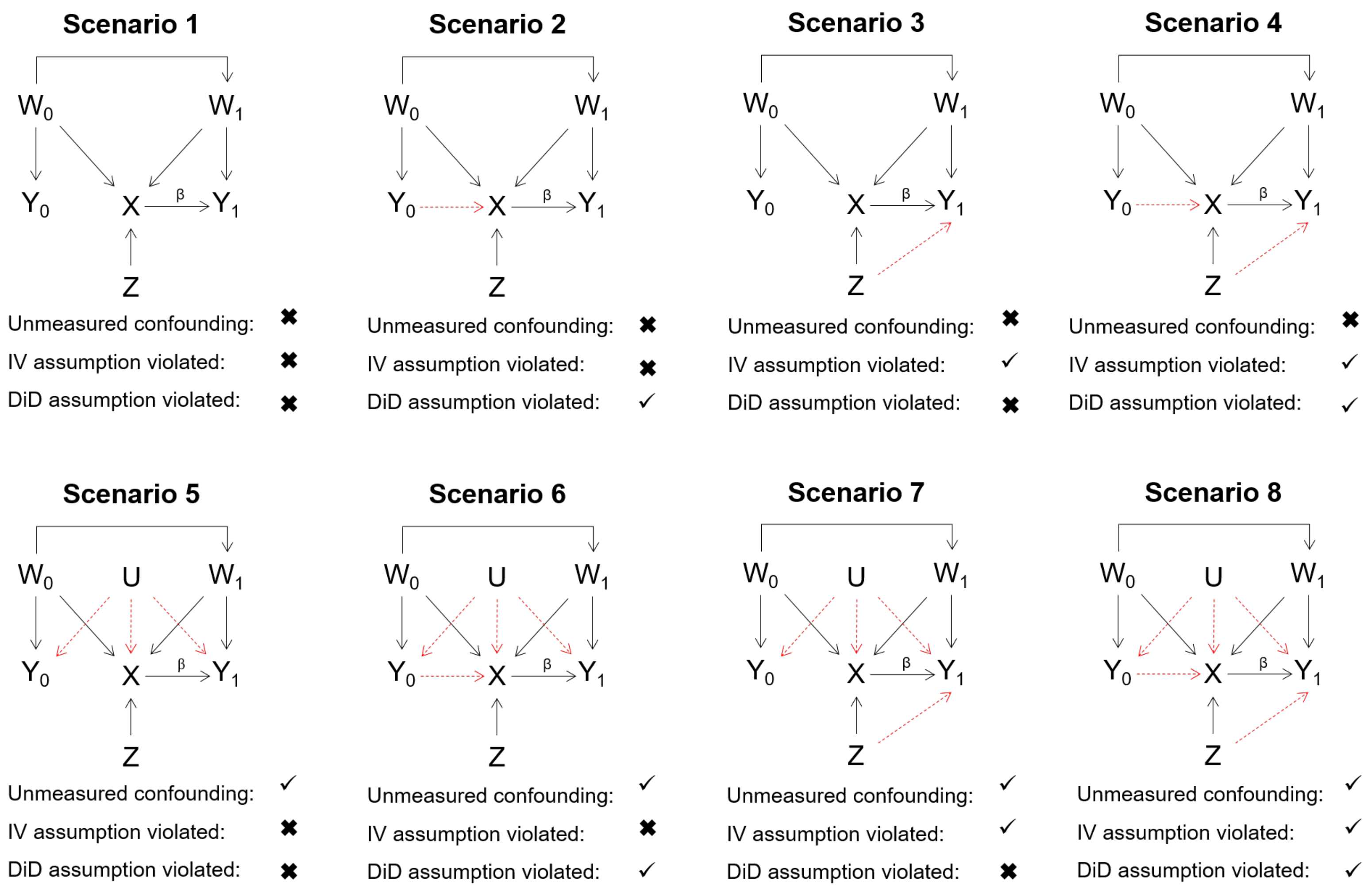

Scenario 1
W_0
W_1
Y_0
X
β
Y_1
Z
Unmeasured confounding: ✖
IV assumption violated: ✖
DiD assumption violated: ✖
Scenario 2
W_0
W_1
Y_0
X
β
Y_1
Z
Unmeasured confounding: ✖
IV assumption violated: ✖
DiD assumption violated: ✓
Scenario 3
W_0
W_1
Y_0
X
β
Y_1
Z
Unmeasured confounding: ✖
IV assumption violated: ✓
DiD assumption violated: ✖
Scenario 4
W_0
W_1
Y_0
X
β
Y_1
Z
Unmeasured confounding: ✖
IV assumption violated: ✓
DiD assumption violated: ✓
Scenario 5
W_0
U
W_1
Y_0
X
β
Y_1
Z
Unmeasured confounding: ✓
IV assumption violated: ✖
DiD assumption violated: ✖
Scenario 6
W_0
U
W_1
Y_0
X
β
Y_1
Z
Unmeasured confounding: ✓
IV assumption violated: ✖
DiD assumption violated: ✓
Scenario 7
W_0
U
W_1
Y_0
X
β
Y_1
Z
Unmeasured confounding: ✓
IV assumption violated: ✓
DiD assumption violated: ✖
Scenario 8
W_0
U
W_1
Y_0
X
β
Y_1
Z
Unmeasured confounding: ✓
IV assumption violated: ✓
DiD assumption violated: ✓

Figure 3.2: DAGs representing the scenarios of the simulation.

3.4.2 Simulation results

Simulation results are summarized for all 8 scenarios in Table 3.2. Specifically, we use the 1000 CaT, IV, CF and DiD estimates to calculate the: bias, and mean squared error (MSE); the mean empirical standard error (SE) arising directly from the model fits; coverage rate of 95% confidence intervals (CI) and the type 1 error (T1E) rate when rejecting the null hypothesis of no causal effect at the 5% significance level. In order to assess the type 1 error simulation calculations were executed with $\beta = 0$. The most efficient and unbiased method for each scenario is highlighted in Table 3.2 in **bold**. Figure 3.3 shows the distribution of the CaT, IV, CF and DiD estimates over all simulation runs. Additional simulation results including the Monte Carlo standard error estimates of the performance measures as described in Morris et al. [219], are given in Appendix 3.3.

For scenario 1, we confirm that the CaT, IV, CF and DiD estimates are all unbiased for the hypothetical estimand β, and the CaT estimate is most efficient. The coverage and T1E rates for all estimates are close to their nominal levels. In scenario 2, the DiD estimate is systematically biased and consequently has poor coverage and T1E. Since a non-zero $\text{Y}_0 - \text{X}$ relationship does not affect the CaT, IV or CF approaches, they estimate the treatment effect without bias, with the CaT being the most efficient. In scenario 3, which was intended to showcase the impact of IV2 assumption violation only, the IV and CF estimates are biased and their coverage/T1E rates are also adversely affected. The diagram for scenario 3 in Figure 3.2 reveals that due to the direct effect of Z on Y_1, Z is an additional confounder, which must be included in the regression models for the CaT and DiD estimates. The results for the CaT and DiD estimates therefore include Z as measured confounder. Again, the CaT estimate is the most efficient in this scenario. In scenario 4 both the IV and DiD assumptions are violated and Z becomes a measured confounder for Y_1 again. Results of the CaT and DiD estimates additionally adjusted for Z as a confounder show that only CaT remains unbiased in this scenario.

For scenario 5 to 8, unmeasured confounding was implemented. Consequently, the CaT estimate is biased and displays lower coverage and higher T1E compared to the IV, CF and DiD estimates, as expected. Comparing the latter two, the DiD estimate is the most efficient unbiased effect measure in scenario 5. When the DiD1 assumption is also violated in scenario 6, its estimate is again biased and shows very low coverage and high T1E rates. Only the IV and CF estimates remain unbiased. Similar to scenario 4, in scenario 7 the DiD estimate will be only biased if Z is not included in the model. The CaT estimate on the other hand remains biased due to the unmeasured confounding. In scenario 8, the identifying assumptions of all three methods are violated. Consequently, none of the methods are able to estimate the treatment effect without bias. This scenario may very well represent the reality of a given analysis setting. For this reason in Section 3.5 we discuss an extension of the standard Instrumental Variable method that can give consistent estimates for the causal effect under a different set of assumptions.

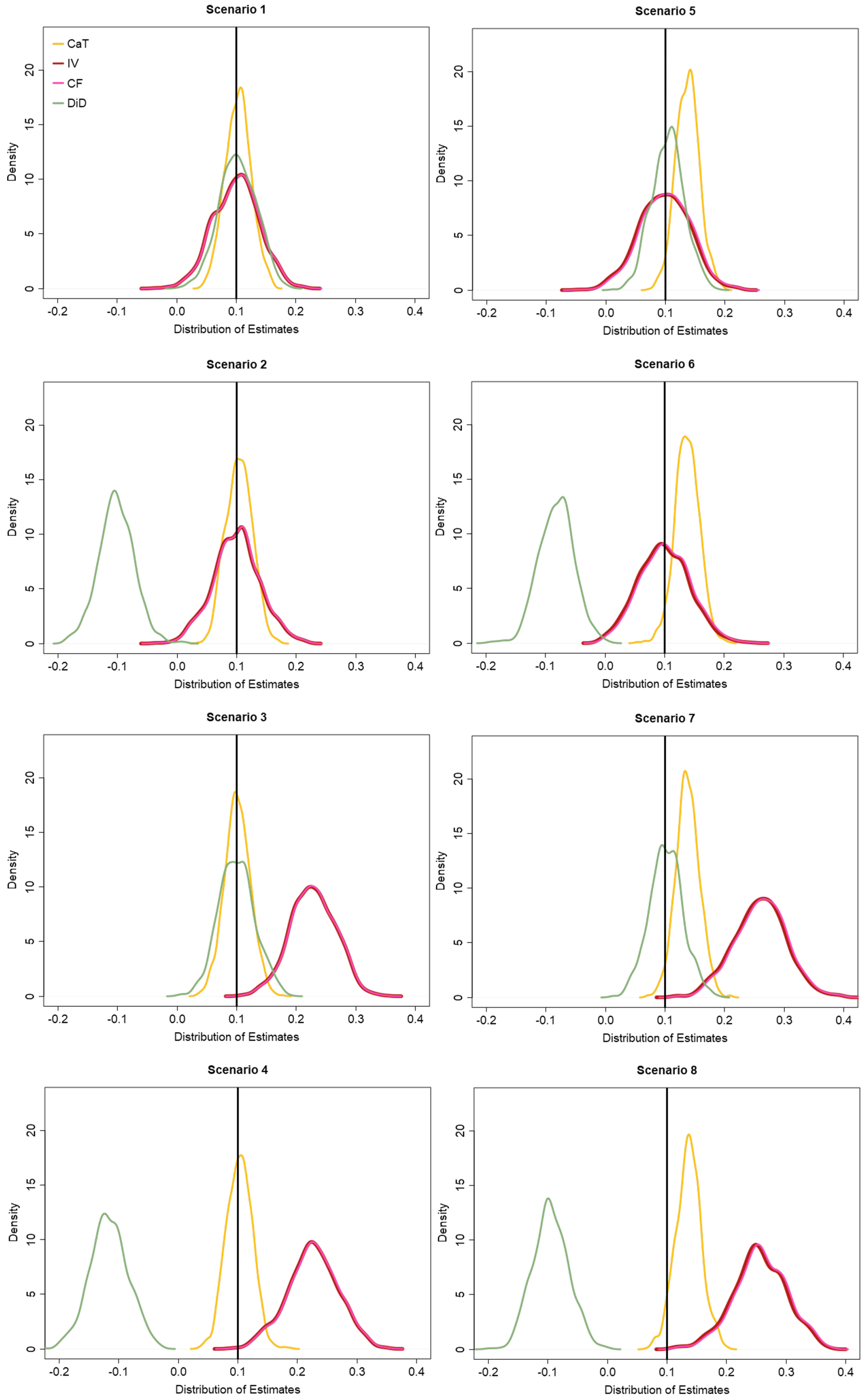


Figure 3.3: Distribution of estimation results for the CaT, IV, CF and DiD method for all simulation scenarios.

		CaT	IV	CF	DiD
Scenario 1	Bias	**0.313**	-0.0117	0.0616	0.251
	SE	**0.0685**	0.121	0.121	0.0988
	MSE	**0.0478**	0.146	0.146	0.0981
	Coverage	**94.1**	95.2	95	93.4
	T1E	**6**	5.7	5.5	5
Scenario 2	Bias	**0.411**	-0.0345	0.0951	-20.2
	SE	**0.0697**	0.125	0.125	0.0956
	MSE	**0.0502**	0.156	0.157	4.17
	Coverage	**95**	95.2	95.2	0
	T1E	**5.3**	4.8	4.8	100
Scenario 3	Bias	**0.134**	12.8	12.9	-0.002
	SE	**0.0687**	0.121	0.121	0.0958
	MSE	**0.0473**	1.78	1.8	0.0916
	Coverage	**94.2**	8.1	8	94.3
	T1E	**4.6**	94.9	94.8	6.3
Scenario 4	Bias	**0.261**	12.6	12.7	-21.6
	SE	**0.0685**	0.135	0.135	0.102
	MSE	**0.0475**	1.77	1.8	4.78
	Coverage	**96.4**	13.9	12.9	0
	T1E	**4.7**	88.3	88.4	100
Scenario 5	Bias	3.72	-0.169	0.111	**0.436**
	SE	0.0625	0.134	0.133	**0.0868**
	MSE	0.177	0.178	0.177	**0.0771**
	Coverage	53.4	95.2	94.7	**95**
	T1E	45.4	5.9	5.9	**4.8**
Scenario 6	Bias	3.77	**-0.16**	**0.184**	-18.2
	SE	0.0649	**0.135**	**0.135**	0.0935
	MSE	0.184	**0.182**	**0.182**	3.39
	Coverage	55.4	**96.4**	**96**	0
	T1E	38.5	**5.6**	**5.6**	100

		CaT	IV	CF	DiD
Scenario 7	Bias	3.8	15.9	16.2	**0.205**
	SE	0.0631	0.14	0.14	**0.0894**
	MSE	0.184	2.72	2.82	**0.0803**
	Coverage	55.5	6.2	5.2	**94.6**
	T1E	45.3	96.5	96.8	**5.1**
Scenario 8	Bias	3.6	15.5	15.9	-19.5
	SE	0.0682	0.141	0.141	0.0958
	MSE	0.176	2.61	2.72	3.9
	Coverage	60.4	7.4	6.5	0
	T1E	39.4	94.8	95.1	100

Table 3.2: Bias, standard errors (SE) and mean squared error (MSE) (all × 100); coverage and type 1 error (T1E) rate (both expressed as a percentage based on a 95% confidence interval and 5% significance threshold) for the estimates CaT, IV, CF and DiD and for all scenarios.

3.4.3 Similarity statistic performance

As proof of concept for the Q_e statistic explained in Section 3.3, we repeat simulation scenario 1, 2, 3 and 5 with 500 simulation runs. In each simulation run, Q_e is calculated using 500 non-parametric bootstraps. Table 3.3 shows the rejection rates (in %) when testing if the CaT, CF and DiD estimates are similar at the 5% level using all three pairwise comparisons.

	e	H_0 rejected (%)	95% CI
Scenario 1	CaT, CF	7.4	5.11; 9.7
	CaT, DiD	6.4	4.25; 8.55
	CF, DiD	6.4	4.25; 8.55
Scenario 2	CaT, CF	6.0	3.92; 8.08
	CaT, DiD	100.0	100; 100
	CF, DiD	100.0	100; 100
Scenario 3	CaT, CF	96.2	94.52; 97.88
	CaT, DiD	4.4	2.6; 6.2
	CF, DiD	87.8	84.93; 90.67
Scenario 5	CaT, CF	18.2	14.82; 21.58
	CaT, DiD	40.8	36.49; 45.11
	CF, DiD	4.0	2.28; 5.72

Table 3.3: Rejection rates and 95% confidence intervals (in %). Results are shown for all pairwise combinations of the estimates.

In Scenario 1 all three estimators target the same true estimand β. Although we see a small degree of type 1 error inflation, rejection rates for each test are reassuringly low. In scenario 2 the data is generated without unmeasured confounding, but the DiD1 assumption is violated. Therefore the DiD estimator targets a distinct estimand from the CaT or CF approaches, which themselves target the same estimand. This is indeed reflected by the test results, where we observe 100% power to detect a difference between the DiD estimate and either the CaT or CF estimates and a low power of 6% to distinguish the CaT and CF estimates themselves. In Scenario 3 (where only the CaT and DiD estimates are truly similar) and Scenario 5 (where only CF and DiD estimates are truly similar) the Q_e statistic exhibits comparable performance. In scenario 5 we can only reliably detect that the DiD and CF estimates are similar but we do not have enough power to detect the differences with the CaT estimate. This might be because the bias due the violation of the NUC assumption is relatively small as well as the true difference in estimates.

3.5 The prior outcome augmented Instrumental Variable method

In this section we introduce the prior outcome augmented Instrumental Variable (POA-IV) estimate which aims to overcome the limitations of the DiD and standard IV estimate by leveraging data from both the prior and study period. Specifically, we look to leverage an interaction between the prior outcome Y_0 and the original IV Z to form a new IV. This general technique to use interaction terms has been successfully applied in several different contexts in recent years. For example to disentangle direct and indirect effects in a mediation analysis [220], to allow for violation of the homogeneity assumption in a non-adherence affected RCT [206], and to adjust for bias due to pleiotropy in Mendelian randomization [221]. The POA-IV estimates follows the same idea as the mentioned studies using interaction terms as instruments.

The treatment effect, $\beta_{\mathsf{POA-IV}}$ can be estimated using a slightly modified TSLS approach. In the first stage model, treatment assignment X is regressed on Z and W_1, but also on Y_0 and the interaction term $\mathsf{Y}_0\mathsf{Z}$ as the new IV:

$$\begin{aligned}\mathsf{Logit}(\mathsf{Pr}[\mathsf{X}=1|\mathsf{Z},\mathbf{W_1},\mathsf{Y}_0]) = \beta_{\mathsf{X},0} + \beta_{\mathsf{X},\mathsf{Z}}\mathsf{Z} + \beta_{\mathbf{X},\mathbf{W_1}}\mathbf{W_1} + \\ \beta_{\mathsf{X},\mathsf{Y}_0}\mathsf{Y}_0 + \beta_{\mathsf{X},\mathsf{Y}_0\mathsf{Z}}\mathsf{Y}_0\mathsf{Z}.\end{aligned} \tag{3.10}$$

By including Y_0 in the first stage model, the estimate acknowledges that the treatment decision can be affected by previously measured outcomes such as drug specific adverse events or other outcomes in the prior period. Fitted values from regression model (3.10), $\hat{\mathsf{X}}$, are then used in the second stage model:

$$\mathsf{E}[\mathsf{Y}_1|\hat{\mathsf{X}},\mathbf{W_1},\mathsf{Y}_0,\mathsf{Z}] = \beta_{\mathsf{Y}_1,0} + \beta_{\mathsf{POA-IV}}\hat{\mathsf{X}} + \beta_{\mathbf{Y_1},\mathbf{W_1}}\mathbf{W_1} + \beta_{\mathsf{Y}_1,\mathsf{Z}}\mathsf{Z} + \beta_{\mathsf{Y}_1,\mathsf{Y}_0}\mathsf{Y}_0, \tag{3.11}$$

to furnish a causal estimate for X whilst additionally controlling for any direct effects of Z and Y_0 on Y_1. The interaction term $\beta_{\mathsf{X},\mathsf{Y}_0\mathsf{Z}}$ in model (3.10) would be present in our setting if a provider only shows a prescription preference in situations where the patient has already experienced an event of interest in the prior period or, more generally, if the strength of preference varies across levels of Y_0.

As it serves as a new IV we require that

- POA-IV1: the interaction term $\beta_{\mathrm{X,Y_0Z}}$ is non-zero and strong (in order to avoid weak instrument bias. [220]);
- POA-IV2: Z and Y_0 (and hence $\hat{X}$ in model (3.11)) are independent of $\mathbf{U}$.

To implement the approach using the CF model for binary outcomes (which we refer to as POA-CF), we again fit model (3.10) to the data to give $\hat{X}$. From this we calculate the residual $\hat{\Delta} = X - \hat{X}$, and then fit the second stage regression model

$$\begin{aligned}\mathrm{Logit}(\Pr[Y_1 = 1|X, \mathbf{W_1}, Y_0, Z, \hat{\Delta}]) = \beta_{Y_1,0} + \beta_{\mathrm{POA-CF}}X + \beta_{\mathbf{Y_1,W_1}}\mathbf{W_1} + \\ \beta_{Y_1,Z}Z + \beta_{Y_1,Y_0}Y_0 + (\beta_{Y_1,\hat{\Delta}} + \beta_{Y_1,Z\hat{\Delta}}Z)\hat{\Delta},\end{aligned} \tag{3.12}$$

before estimating the causal effect on the risk difference scale using the `margins()` package [215] as before. The performance of both the POA-CF and equivalent standard IV method (referred to as POA-IV) are explored in the next section.

3.5.1 Simulation study

We now showcase the ability of the POA-IV and POA-CF estimate in comparison to the CaT, IV, CF and DiD estimates under conditions in which the latter three approaches are biased. The simulation is therefore an extension of the simulation described in Section 3.4. The left side of Figure 3.4 clarifies how the data for each scenario is generated. For scenario 1 and 2 of this simulation study the prior outcome Y_0 is generated without unmeasured confounding. Additionally, in scenario 2 Y_0 has a direct effect on the study outcome of interest Y_1. Scenario 3 is the same as scenario 8 of the simulation in Section 3.4. Further information about the data generation models are outlined in Appendix 3.4.

3.5.2 Simulation results

The results of scenario 1 in Table 3.4 and the right side of Figure 3.4 show that the POA-IV and POA-CF are able to estimate the true causal risk difference (10%) without bias and similar efficiency to the standard IV estimate. All other methods compared in this simulation exhibit bias. This is also the case for scenario 2 in which Y_0 exerts a direct effect on Y_1. For this scenario Y_0 was included in the

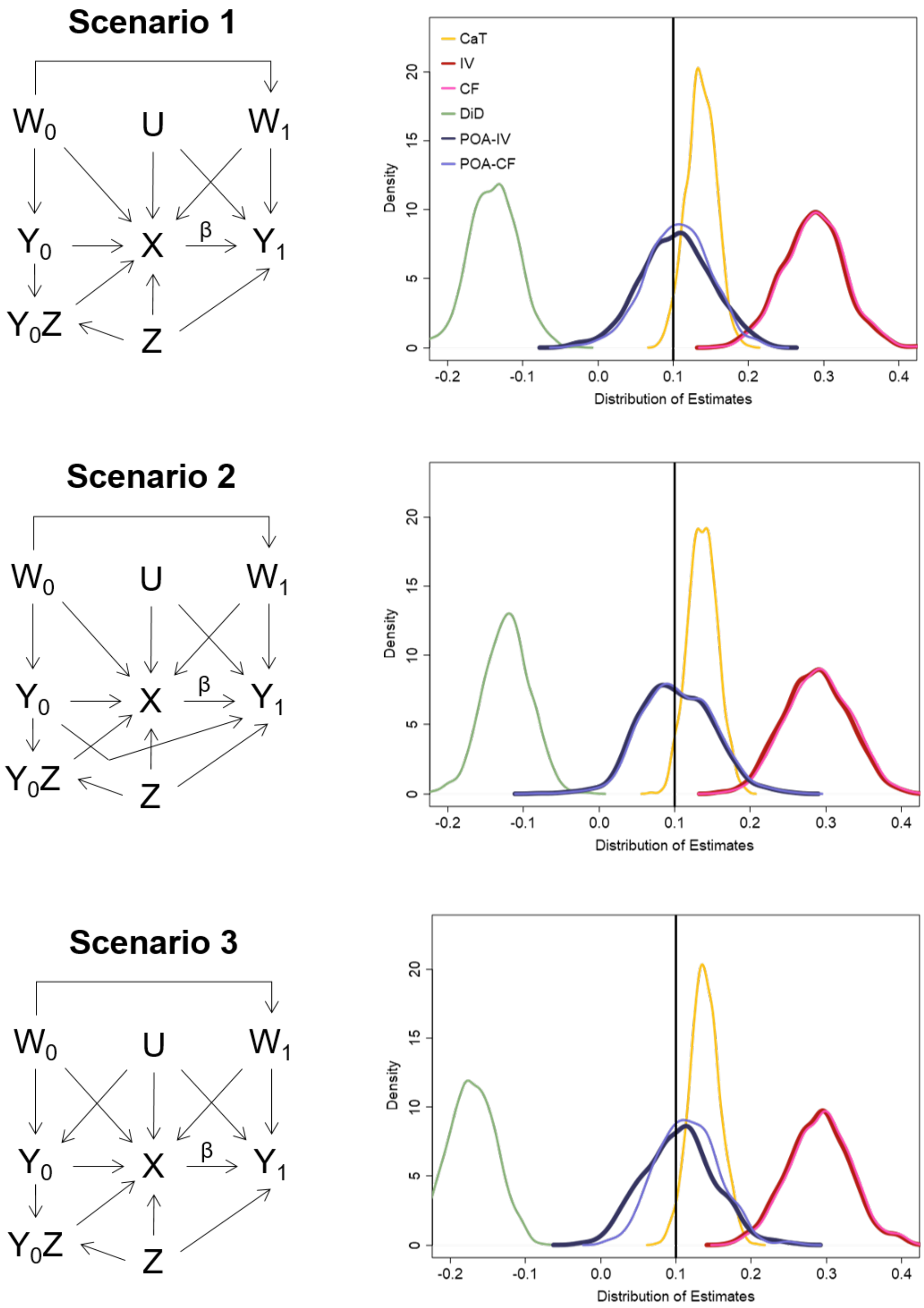


Figure 3.4: Left side: DAG representing the data generation for each scenario of the simulation. Right side: Distribution of the estimation results.

outcome models as measured confounder. Coverage rates of the POA-IV and POA-CF are around 95%. The bias of the IV and CF approach in all scenarios stems from a relatively large effect of Z on Y_1. Scenario 3 of this simulation is the same as scenario 8 described in Section 3.4. All previously applied methods exhibited noticeable bias. As Y_0 is confounded with U in this scenario, POA-IV as well as POA-CF are biased too. From the results of this simulation the bias is much smaller than the bias of CaT, IV, CF and DiD, but it would increase in case of a stronger effect of U on Y_0. Additional information on the Monte Carlo simulation errors are given in Appendix 3.5.

		CaT	IV	CF	DiD	POA-IV	POA-CF
Scenario 1	Bias	3.72	18.6	18.9	-23.9	0.369	0.636
	SE	0.0615	0.129	0.129	0.101	0.15	0.136
	MSE	0.176	3.62	3.73	5.82	0.225	0.188
	Coverage	52.8	0.4	0.4	0	94.4	95.2
	T1E	44.4	99.4	99.4	100	5.9	5.9
Scenario 2	Bias	3.71	18.7	19.1	-22.2	0.219	0.527
	SE	0.0613	0.135	0.135	0.0974	0.15	0.15
	MSE	0.175	3.69	3.85	5.02	0.227	0.228
	Coverage	53.6	0.6	0.5	0	95.7	95.6
	T1E	45.7	99.3	99.4	100	4	4.2
Scenario 3	Bias	3.85	19	19.3	-27.1	0.428	1.59
	SE	0.0629	0.131	0.131	0.0998	0.149	0.131
	MSE	0.187	3.77	3.9	7.45	0.225	0.196
	Coverage	50.7	0.3	0.3	0	95.5	92.8
	T1E	48.9	99.7	99.7	100	6	6.2

Table 3.4: Bias, standard errors (SE) and mean squared error (MSE) (all × 100); coverage and type 1 error (T1E) rate (both expressed as a percentage based on a 95% confidence interval and 5% significance threshold) for the estimates CaT, IV, CF, DiD, POA-IV and POA-CF and for all scenarios.

3.6 Application to type 2 diabetes patients in Clinical Practice Research Datalink

In addition to lifestyle modification, treatment for type 2 diabetes (T2D) primarily focuses on the management of blood glucose, with different glucose-lowering oral agents available. Metformin (MFN) is recommended as first-line medical therapy by major T2D clinical guidelines [124, 222], but if glucose control deteriorates, additional second-line or further treatments are prescribed. Sodium-glucose Cotransporter-2 Inhibitors (SGLT2i) and Dipeptidyl peptidase-4 Inhibitors (DPP4i) are two widely used second-line medication classes in the UK and US [127, 223], and there is considerable interest in using observational data to establish the comparative benefits and risks of the two therapies in 'real-world' settings and for a broad spectrum of patients. [224] Whilst SGLT2is have some benefits beyond blood sugar lowering (including reducing the risk of cardiovascular disease) they may be associated with increased risk for genital infection. [124]

We used routine data from the Clinical Practice Research Datalink (CPRD Gold, download July 2019) to examine the risk of genital infections for people with T2D initiating SGLT2i ($N_{Tx} = 1966$) compared to DPP4i ($N_{Ct} = 4033$) during 2016-2019 as second-line treatment after MFN. [200, 203] CPRD is a rich source of primary care data for observational health research. This database includes approximately 6.9% of the UK population and patients are considered to be representative with regard to age, sex and ethnicity. [200] Studying the efficacy and tolerability with routine practice data makes it possible to understand the risks and benefits of medication use in a large and truly representative population in contrast to clinical trials which are performed on a population restricted by factors such as age or diabetes severity. [225, 226]

All individuals in the study cohort initiated MFN as first-line treatment and have not been prescribed insulin over the complete follow-up time. Additionally, only individuals who initiated DPP4i or SGLT2i as second-line treatment were included in the analysis. The prior period is observed from start of the initiation of MFN until just before the start of the second-line treatment (SGLT2i or DPP4i). The average follow-up time in this period was 3.66 years. Therefore, the study period starts with the initiation of the second-line treatment until one of the following-censoring reasons: end of follow-up data (30th of June 2019), discontinuation of second-line treatment or start of the other comparison treatment (e.g. individual started to take SGLT2i as second-line treatment and added DPP4i at a later point in time). The average follow-up time of the study period was 1.44 years.

The baseline characteristics of the cohort are summarized in Table 3.5, for the two periods before and after initiation of second-line treatment. In the prior period 151 (2.5%) genital infections are recorded, 45 (2.3%) and 106 (2.6%) for people on SGTL2i and DPP4i respectively. In the study period 139 (2.3%) people experience an infection, 96 (4.9%) on SGLT2i and 43 (1.1%) on DPP4i. Genital infection is therefore a rare outcome. Data was also extracted on patients' general practice membership, in order to use it as an IV within the standard and prior outcome augmented IV approach.

The outcome, defined as $\geq$ 1 genital infection in a given period, was coded as a binary variable and modelled using logistic regression. Causal estimates are reported on the risk difference scale (in %) as described in Section 3.3.4.

For our analysis we applied the six causal estimation strategies introduced in Sections 3.3 and 3.5 to estimate the population averaged effect of taking SGLT2i versus DPP4i on infection risk. Additionally, we applied the CaT estimator on a propensity score matched dataset (PSM) using the R package `MatchIt()` with 1-1 nearest neighbour matching with replacement. [227] Approximately two-thirds of the data was matched. The balance diagnostic statistics are summarized with a love-plot in Appendix 3.6 and show that the matching procedure has improved the balance of the treatment groups. This plot also gives a list of the variables which was used for the matching procedure. Furthermore, the CaT and DiD method were applied including Z as measured confounder to avoid bias in case the exclusion restriction of the IV method is not met, which cannot be verified with the data at hand. [70] Y_0 was included as a measured confounder in all models for Y_1 as it has been found in previous studies that prior infections are associated with the risk of experiencing an infection in the study period. [157] The β_{CaT} estimate was obtained from a multivariable logistic regression adjusted for all baseline characteristics measured at second-line treatment initiation, as listed in Table 3.5. The β_{DiD} estimate was obtained using logistic difference-in-difference regression and also adjusted for the baseline characteristics at initiation of first- and second-line treatment. Standard IV, CF and the prior outcome augmented IV approaches were fitted to the data using the methods previously described, with adjustment for the same set of baseline covariates in the first and second stage models.

3.6.1 Construction of the Instrumental Variable

As prescription prevalence of both drug classes increased dramatically after 2015 and regional differences in prescribing patterns in the UK exist [126, 127], the IV Z constructed for this analysis aims to convey information about providers' preference to prescribe SGLT2i over DPP4i. Preference-based IVs have been proposed when it is assumed that providers prescription preference varies or a substantial variation in practice pattern can be observed. [92, 80] Hence, for the estimation of β_{IV}, β_{CF}, $\beta_{\text{POA}-\text{IV}}$ and $\beta_{\text{POA}-\text{CF}}$, we constructed a binary IV for patients treated by each respective provider as proposed by Brookhart et al. [92]. As the prescription preference is unobserved, a proxy variable is constructed using the observed prescription behaviour of each provider. Further information about this proxy design can be found for example in Davies et al. [91] or Widding et al. [59]. The healthcare provider is assumed to have a preference to prescribe SGLT2i over DPP4i depending on the most recent prescription at each point in time. The patient data of the first patient treated within each provider was excluded from the analysis as the IV could not be calculated for this patient. As SGLT2i and DPP4i are newer drug classes and started to be prescribed as second-line treatment more often after 2014 [127, 223], we allowed for an initial period in which preference could develop. Therefore, data of individuals initiating second-line treatment from 2016 onwards is analysed. Furthermore, we use an IV which makes it possible to account for changes in prescription preferences. [92] A similar approach of using clinical commissioning group prescribing history as preference-based IV has been proposed to evaluate T2D treatment by Bidulka et al. [83].

	Prior period		Study period	
Variable	DPP4i	SGLT2i	DPP4i	SGLT2i
HbA1c (mmol/mol)	70.57 (19.3)	72.43 (19.75)	69.85 (15.27)	73.95 (16.01)
BMI (kg/m^2)	33.24 (6.33)	36.35 (6.8)	32.45 (6.38)	35.8 (6.7)
eGFR (ml/min/1.73m^2)	85.03 (19.74)	92.7 (18.2)	83.18 (22.86)	93.08 (18.75)
Age (years)	61.09 (10.97)	55.26 (9)	65.02 (11.58)	58.61 (9.24)
T2D duration (years)	2.3 (3.01)	1.84 (2.59)	6.21 (4.36)	5.17 (3.6)
Gender				
female	41.04%	60.99%		
male	58.95%	39.01%		
Prescription year				
2016			30.17%	18.81%
2017			30.83%	31.56%
2018			28.61%	34.42%
2019			10.39%	15.21%

Table 3.5: Baseline data on CPRD T2D cohort for prior and study period and for patients on DPP4i ($N_{Ct} = 4033$) or SGLT2i ($N_{Tx} = 1966$) as second-line treatment. Values are shown in mean (standard deviation) unless otherwise stated.

3.6.2 Results

The results of the causal analysis are given in Table 3.6 and Figure 3.5. All methods estimate a positive causal effect suggesting that genital infection risk is higher if all people initiated SGLT2i compared to DPP4i. The POA-IV and POA-CF causal estimates are not significantly different from zero at or below the 5% significance threshold. The POA-IV and POA-CF estimate the causal effect with large uncertainty compared to all other approaches and consequently its 95% confidence interval crosses the null. Although the POA-IV and POA-CF estimate can deal with a direct effect of the prior outcome Y_0 on future treatment X, it assumes no unmeasured confounding between Y_0 and X. Including Z in the DiD and CaT model does not result in a big change of the estimation results in this application case study.

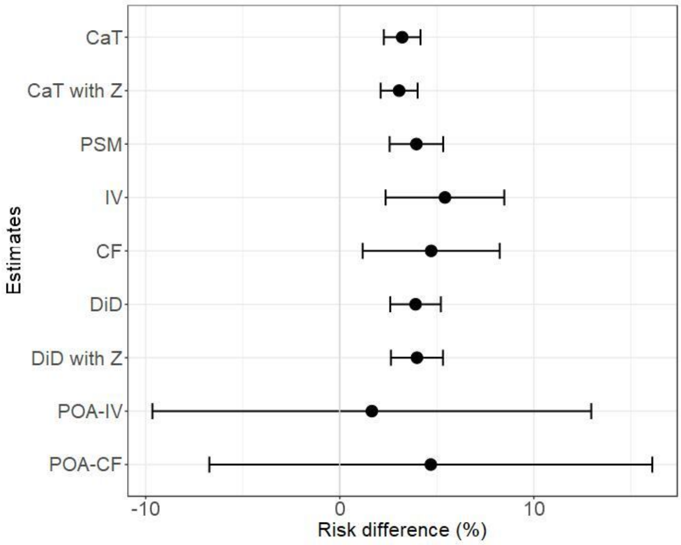


Figure 3.5: Estimated treatment effect for all estimates and their 95% confidence intervals.

Method	Estimate	95% CI	SE	p-value
CaT	3.22	2.27, 4.16	0.49	3.10×10^{-14}
CaT with Z	3.06	2.1, 4.01	0.49	2.08×10^{-12}
PSM	3.95	2.57, 5.33	0.72	5.72×10^{-10}
IV	5.42	2.36, 8.48	2.42	0.0003
CF	4.71	1.18, 8.24	2.77	0.008
DiD	3.91	2.6, 5.21	0.66	7.46×10^{-10}
DiD with Z	3.98	2.64, 5.32	0.67	1.16×10^{-9}
POA-IV	1.65	-9.65, 12.96	6.81	0.77
POA-CF	4.69	-6.72, 16.11	6.79	0.42

Table 3.6: Estimation results on risk difference scale (in %), standard error, and p-value of the estimated treatment effect.

Table 3.7 summarized the strength of the IVs measured with the F-statistic of the coefficient of each IV from the first stage regressions of each respective method. [77] IV and CF as well as POA-IV and POA-CF use the same first stage regression model and the results of the IV strength are therefore summarized in the same row. The instrument strength of Z for the IV and CF approach is strong with F-statistic values greater than 10 but Y_0Z does not seem to be a strong instrument for the POA-IV and POA-CF approach. This helps to understand the poor results of the two methods which estimate the treatment effect with much higher uncertainty than all other methods applied in this study. Furthermore, it is plausible that Y_0 is confounded by U which will also lead to biased results for the POA-IV and POA-CF.

Models	Instrument	F-statistic
IV and CF	Z	345.42
POA-IV and POA-CF	Y_0Z	0.61

Table 3.7: Strength of the Instrumental Variables measured with the F-statistic of Z (for IV and CF) and Y_0Z (POA-IV and POA-CF) from the corresponding first stage regression models.

3.6.3 Results of the similarity statistic

We now apply our Q_e statistic analysis to the set of estimators to assess their similarity. Table 3.8 shows the Q_e statistic for a selection of estimator sets. The pairwise correlation of all estimates calculated over 500 bootstrap samples is summarized in Figure 3.10 in Appendix 3.7. Interestingly, the test statistic for the closely related CaT and PSM estimates as well as for the POA-IV and POA-CF estimates reveal they are not sufficiently similar even though their values are very close. This is explained by their very high correlation, which Q_e adjusts for. IV and CF are identified as sufficiently similar even after accounting for their correlation. All other selected combinations which do not include CaT or PSM and POA-IV or POA-CF together show that the estimates are statistically similar.

e	$\mathrm{Q_e}$ statistic	χ^2 value (df)	p-value	Test decision
CaT, PSM	9.721	3.841 (1)	0.0018	not similar
IV, CF	0.093	3.841 (1)	0.761	similar
POA-IV, POA-CF	9.086	3.841 (1)	0.0026	not similar
CaT, CF, DiD, POA-CF	2.848	7.815 (3)	0.4157	similar
PSM, CF, DiD, POA-CF	0.343	7.815 (3)	0.9517	similar
CaT, IV, DiD, POA-IV	4.101	7.815 (3)	0.2508	similar
PSM, IV, DiD, POA-IV	1.367	7.815 (3)	0.7132	similar

Table 3.8: Test results of the heterogeneity test with $\mathrm{Q_e}$ *statistic and 95% confidence.*

3.7 Summary and conclusion

In this paper we propose a framework for the application of several causal inference methods to assess the comparative effectiveness of two treatments in observational data. This included 'standard' confounder adjustment approaches such as multivariable regression and propensity score matching, difference-in-differences and IV estimation. The assumptions of each approach were described, and a simulation was used to assess the impact of violating necessary assumptions on the estimators' performance. Building on the work of Bowden et. al. [11], we proposed the use of a similarity statistic to formally assess the level of agreement between sets of estimates that can account for their underlying correlation. We hope this statistic could be a useful tool when attempting to triangulate findings from a set of distinct causal estimation strategies going forward.

We illustrated the application of these methods using routinely collected data on people with T2D, to assess the relative safety of SGLT2i compared to DPP4i as second-line therapies on the risk of genital infection. Our heterogeneity analysis showed good agreement between all causal estimates except the PSM/ CaT and POA-IV/POA-CF approaches. In future work, we plan to apply the same causal framework to model alternative T2D outcomes such as HbA1c and other clinically important adverse events. We also plan to extend the approach to fit alternative models that allow for causal effect heterogeneity, so that they may be used in personalised medicine. [228] Furthermore, the applied analysis showcased how triangulating estimation results with different methods can help to identify implausible results and give further insight in possible assumption violations. Additional work still needs to be done about which estimation results final reports should focus on. Possible strategies could be to only focus on similar estimates or to combine the results.

We proposed the use of the POA-IV/ POA-CF method which is able to leverage an interaction between the prior period and the IV accounting for a possible direct effect of the IV on the outcome and a direct effect of previous outcome events on the treatment decision. Our simulations show that this approach is robust and leads to reliable results in scenarios in which key assumptions of the DiD and the IV approaches are violated, as long as the prior outcome-future treatment relationship does not suffer from unmeasured confounding. Furthermore, our simulation in Section 3.5.1 showed that POA-IV and POA-CF were less biased than CaT, DiD, IV and CF even if Y_0 was confounded by U. As future work we hope to better understand when this will be the case.

As further research, we plan to develop a rigorous hierarchical testing procedure for performing a similarity analysis across an arbitrary number of estimates, whilst controlling the overall family wise error rate. Another approach for combining IV and DiD approaches has been proposed by Ye et al. [229]. The ‘instrumented DiD’ purports to offer robustness to time-varying unmeasured confounding and therefore offers utility as an additional estimator within a triangulation analysis.

Acknowledgements

This article is based in part on data from the Clinical Practice Research Datalink obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. CPRD data is provided by patients and collected by the NHS as part of their care and support. This study was supported by the National Institute for Health and Care Research Exeter Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Funding

Approval for the study was granted by the CPRD Independent Scientific Advisory Committee (ISAC 13_177R). LMG, JB and JMD are supported by the UK Research and Innovation Expanding Excellence in England fund. BMS and JMD receive funding from MRC and MASTERMIND. This research was funded by the Medical Research Council (UK) (MR/N00633X/1).

Conflict of interest

JB is a part time employee of Novo Nordisk. This project is unrelated to his work for the company.

Data availability statement

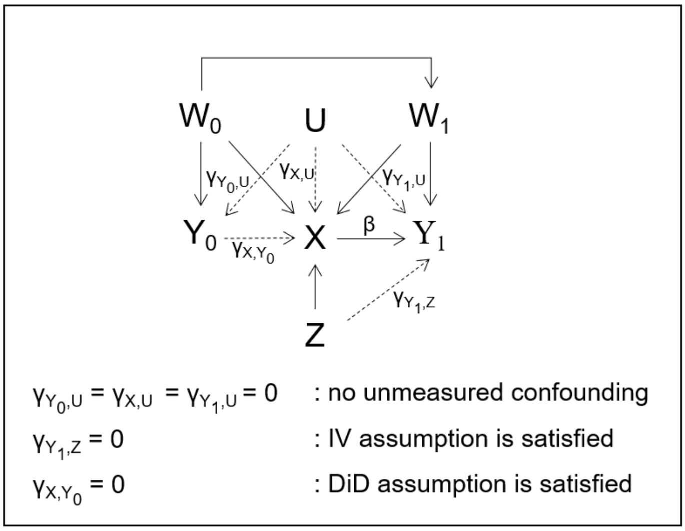
Data from CPRD is available to all researchers following successful application to the ISAC. Source code for this research for all simulations and the application case study in this paper is available at `https://github.com/GuedemannLaura/POA-IV`.

3.8 Appendices

Appendix 3.1 Data generation models of the first simulation study

Data for the first simulation described in Section 3.4 was generated under the models listed below. The DAG in Figure 3.6 visualizes the data structure of the simulation explained in Section 3.4 as well as the mechanisms with which the simulation scenarios are implemented.

$$
\begin{aligned}
\beta &= 0.1\\
Z_{ij} &\sim \mathrm{Bern}(0.5)\\
W_{0i} &\sim N(0,1)\\
W_{1i} &= \gamma_{W_1,W_0} W_{0i} + \gamma_{W_1,\varepsilon}\varepsilon_{W_1 i}\\
\varepsilon_{W_1 i} &\sim N(0,1)\\
U_i &\sim N(0,1)\\
Y_{0i} &\sim \mathrm{Bern}(\gamma_{Y_0,0} + \gamma_{Y_0,U} U_i + \gamma_{Y_0,W_0} W_{0i})\\
X_i &\sim \mathrm{Bern}(\gamma_{X,0} + \gamma_{X,Z} Z_{ij} + \gamma_{X,U} U_i + \gamma_{X,W_0} W_{0i} + \gamma_{X,W_1} W_{1i} + \gamma_{X,Y_0} Y_{0i})\\
Y_{1i} &\sim \mathrm{Bern}(\gamma_{Y_1,0} + \gamma_{Y_1,U} U_i + \beta X_i + \gamma_{Y_1,W_1} W_{1i} + \gamma_{Y_1,Z} Z_{ij})
\end{aligned}
$$



$\gamma_{Y_0,U} = \gamma_{X,U} = \gamma_{Y_1,U} = 0$: no unmeasured confounding

$\gamma_{Y_1,Z} = 0$: IV assumption is satisfied

$\gamma_{X,Y_0} = 0$: DiD assumption is satisfied

Figure 3.6: Causal DAG consistent with the data generation of the simulation outlined in Section 3.4.

Appendix 3.2 Proof for DiD assumptions

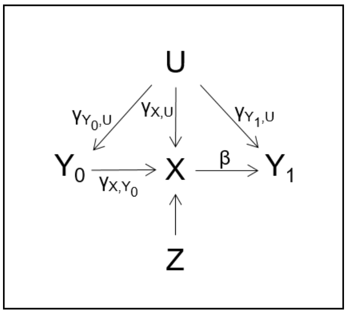


Figure 3.7: A simplified parameterised causal diagram to accompany the DiD proof argument below.

The parameterised causal diagram in Figure 3.7 indicates a similar structure to that in Section 3.2 but without measured confounders W_0 and W_1 for simplification. Removing the individual subscript i for convenience, assume the following models for Y_0, X and Y_1:

$$\begin{aligned} Y_0 &= \gamma_{Y_0,U}U + \epsilon_{Y0} && (3.13)\\ X &= \gamma_{X,U}U + \gamma_{X,Y_0}Y_0 + \epsilon_X && (3.14)\\ Y_1 &= \beta X + \gamma_{Y_1,U}U + \epsilon_{Y1}, && (3.15) \end{aligned}$$

where β represents the causal effect that DiD is attempting to estimate. The estimand targeted by a regression of Y_1 on X is therefore

$$\frac{\text{Cov}(Y_1, X)}{\text{Var}(X)} = \frac{\beta\text{Var}(X) + \gamma_{Y1,U}\text{Cov}(X, U)}{\text{Var}(X)}, \tag{3.16}$$

and the estimand targeted by a regression of Y_0 on X is therefore

$$\frac{\text{Cov}(Y_0, X)}{\text{Var}(X)} = \frac{\text{Cov}(\gamma_{Y_0,U}U + \epsilon_{Y0}, X)}{\text{Var}(X)}. \tag{3.17}$$

Putting (3.16) and (3.17) together, DiD estimand can be written as

$$\frac{\text{Cov}(Y_1, X)}{\text{Var}(X)} - \frac{\text{Cov}(Y_0, X)}{\text{Var}(X)} = \beta + (\gamma_{Y_1,U} - \gamma_{Y_0,U})\frac{\text{Cov}(U, X)}{\text{Var}(X)} - \gamma_{X,Y_0}\frac{\text{Var}(\epsilon_{Y_0})}{\text{Var}(X)}. \tag{3.18}$$

From (3.18) we see that that the DiD estimand is equal to β when $\gamma_{Y_0,U} = \gamma_{Y_1,U}$ (DiD2 assumption) and either γ_{X,Y_0} is zero (DiD1 assumption), or that $\text{Var}(\epsilon_{Y_0}) = 0$

.

Appendix 3.3 Results of Monte Carlo standard errors of the first simulation study

For the simulation of Section 3.4 the Monte Carlo standard errors (MCSE) calculated based on Morris et al. [219]. The results are given in the table below for the performance measures: bias, mean squared error, coverage and type 1 error.

		CaT	IV	CF	DiD
Scenario 1	MCSE(bias)	0.0685	0.1207	0.1209	0.0988
	MCSE(MSE)	0.002	0.0064	0.0064	0.0041
	MCSE(coverage)	0.7451	0.676	0.6892	0.7851
	MCSE(T1E)	0.751	0.7332	0.7209	0.6892
Scenario 2	MCSE(bias)	0.0697	0.1251	0.1252	0.0956
	MCSE(MSE)	0.002	0.0071	0.0072	0.1291
	MCSE(coverage)	0.6892	0.676	0.676	0
	MCSE(T1E)	0.7085	0.676	0.676	0
Scenario 3	MCSE(bias)	0.0687	0.1208	0.1209	0.0958
	MCSE(MSE)	0.0021	0.0522	0.0529	0.004
	MCSE(coverage)	0.7392	0.8628	0.8579	0.7332
	MCSE(T1E)	0.6624	0.6957	0.7021	0.7683
Scenario 4	MCSE(bias)	0.0685	0.1346	0.1345	0.1021
	MCSE(MSE)	0.0022	0.0507	0.0518	0.148
	MCSE(coverage)	0.5891	1.094	1.06	0
	MCSE(T1E)	0.6693	1.0164	1.0126	0
Scenario 5	MCSE(bias)	0.0625	0.1336	0.1332	0.0868
	MCSE(MSE)	0.0047	0.0078	0.0078	0.0033
	MCSE(coverage)	1.5775	0.676	0.7085	0.6892
	MCSE(T1E)	1.5744	0.7451	0.7451	0.676
Scenario 6	MCSE(bias)	0.0649	0.1348	0.1349	0.0935
	MCSE(MSE)	0.005	0.0078	0.0078	0.1045
	MCSE(coverage)	1.5719	0.5891	0.6197	0
	MCSE(T1E)	1.5387	0.7271	0.7271	0

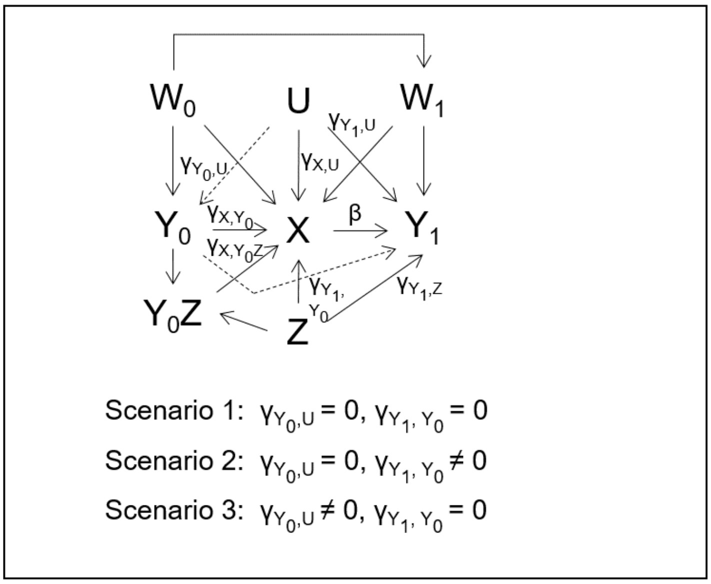
		CaT	IV	CF	DiD
Scenario 7	MCSE(bias)	0.0631	0.1397	0.1399	0.0894
	MCSE(MSE)	0.0049	0.0805	0.0835	0.0036
	MCSE(coverage)	1.5715	0.7626	0.7021	0.7147
	MCSE(T1E)	1.5741	0.5812	0.5566	0.6957
Scenario 8	MCSE(bias)	0.0682	0.1412	0.141	0.0958
	MCSE(MSE)	0.0046	0.0768	0.0803	0.1205
	MCSE(coverage)	1.5466	0.8278	0.7796	0
	MCSE(T1E)	1.5452	0.7021	0.6826	0

Table 3.9: Monte Carlo standard errors (MCSE) of the performance measures of all estimates and all scenarios of the simulation outlined in Section 3.4. All results are multiplied with 100 and rounded to 3 significant figures.

Appendix 3.4 Data generation models of the second simulation study

For the simulation demonstrating the POA-IV and POA-CF estimates the data was generated using the same strategy as for the simulation explained in Section 3.5.1, except for X and Y_1. The data generation models are shown below and Figure 3.8 shows the DAG explaining the mechanisms with which the simulation scenarios are implemented.

$$
\begin{aligned}
\beta &= 0.1 \\
\mathsf{Z}_{\mathsf{ij}} &\sim \mathsf{Bern}(0.5) \\
\mathsf{W}_{\mathsf{0i}} &\sim \mathsf{N}(0,1) \\
\mathsf{W}_{\mathsf{1i}} &= \gamma_{\mathsf{W_1,W_0}}\mathsf{W}_{\mathsf{0i}} + \gamma_{\mathsf{W_1},\varepsilon}\varepsilon_{\mathsf{W_1 i}} \\
\varepsilon_{W_1 i} &\sim N(0,1) \\
\mathsf{U}_{\mathsf{i}} &\sim \mathsf{N}(0,1) \\
\mathsf{Y}_{\mathsf{0i}} &\sim \mathsf{Bern}(\gamma_{\mathsf{Y_0,0}} + \gamma_{\mathsf{Y_0,U}}\mathsf{U}_{\mathsf{i}} + \gamma_{\mathsf{Y_0,W_0}}\mathsf{W}_{\mathsf{0i}}) \\
\mathsf{X}_{\mathsf{i}} &\sim \mathsf{Bern}(\gamma_{\mathsf{X,0}} + \gamma_{\mathsf{X,Z}}\mathsf{Z}_{\mathsf{ij}} + \gamma_{\mathsf{X,U}}\mathsf{U}_{\mathsf{i}} + \gamma_{\mathsf{X,W_0}}\mathsf{W}_{\mathsf{0i}} + \gamma_{\mathsf{X,W_1}}\mathsf{W}_{\mathsf{1i}} + \\
&\qquad \gamma_{\mathsf{X,Y_0}}\mathsf{Y}_{\mathsf{0i}} + \gamma_{\mathsf{X,Y_0Z}} \cdot \mathsf{Z}_{\mathsf{ij}} \cdot \mathsf{Y}_{\mathsf{0i}}) \\
\mathsf{Y}_{\mathsf{1i}} &\sim \mathsf{Bern}(\gamma_{\mathsf{Y_1,0}} + \gamma_{\mathsf{Y_1,U}}\mathsf{U}_{\mathsf{i}} + \beta\mathsf{X}_{\mathsf{i}} + \gamma_{\mathsf{Y_1,W_1}}\mathsf{W}_{\mathsf{1i}} + \gamma_{\mathsf{Y_1,Z}}\mathsf{Z}_{\mathsf{ij}} + \gamma_{\mathsf{Y_1,Y_0}}\mathsf{Y}_{\mathsf{0i}})
\end{aligned}
$$



Scenario 1: $\gamma_{Y_0,U} = 0$, $\gamma_{Y_1,Y_0} = 0$

Scenario 2: $\gamma_{Y_0,U} = 0$, $\gamma_{Y_1,Y_0} \neq 0$

Scenario 3: $\gamma_{Y_0,U} \neq 0$, $\gamma_{Y_1,Y_0} = 0$

Figure 3.8: Causal DAG consistent with the data generation of the simulation outlined in Section 3.5.1

Appendix 3.5 Results of Monte Carlo standard errors of the second simulation study

The Monte Carlo standard errors (MCSE) are calculated based on Morris et al. [219] for the simulation presented in Section 3.5.1. The results are given in the table below for the performance measures: bias, mean squared error, coverage and type 1 error.

		CaT	IV	CF	DiD	POA-IV	POA-CF
Scenario 1	MCSE(bias)	0.0615	0.1287	0.1288	0.1012	0.1496	0.1356
	MCSE(MSE)	0.0047	0.1095	0.113	0.1809	0.0098	0.0083
	MCSE(coverage)	1.5787	0.1996	0.1996	0	0.7271	0.676
	MCSE(T1E)	1.5712	0.2442	0.2442	0	0.7451	0.7451
Scenario 2	MCSE(bias)	0.0613	0.1353	0.1354	0.0974	0.1505	0.1502
	MCSE(MSE)	0.0047	0.1112	0.1162	0.1559	0.0104	0.0105
	MCSE(coverage)	1.5770	0.2442	0.2230	0	0.6415	0.6486
	MCSE(T1E)	1.5753	0.2636	0.2442	0	0.6197	0.6343
Scenario 3	MCSE(bias)	0.0629	0.1313	0.1314	0.0998	0.1493	0.1306
	MCSE(MSE)	0.005	0.1142	0.1181	0.2327	0.0098	0.0079
	MCSE(coverage)	1.581	0.1729	0.1729	0	0.6556	0.8174
	MCSE(T1E)	1.5808	0.1729	0.1729	0	0.751	0.7626

Table 3.10: Monte Carlo standard errors (MCSE) of the performance measures of all estimates and all scenarios of the simulation outlined in Section 3.5.1. All results are multiplied with 100 and rounded to 3 significant figures.

Appendix 3.6 Balance statistic for propensity score matched data

The propensity score matching procedure matched 100% of the 1966 individuals treated with SGLT2i. Therefore, overall 67.43% of all individuals in the data were matched. No records were discarded for the matching procedure. The love plot in Figure 3.9 shows that the matched data improved the balance of groups based on the absolute standardize mean difference. The matching process was employed using the baseline characteristics shown in the figure measured at first-line treatment initiation and years of second-line treatment initiation as this covariate has no effect on the treatment effect.

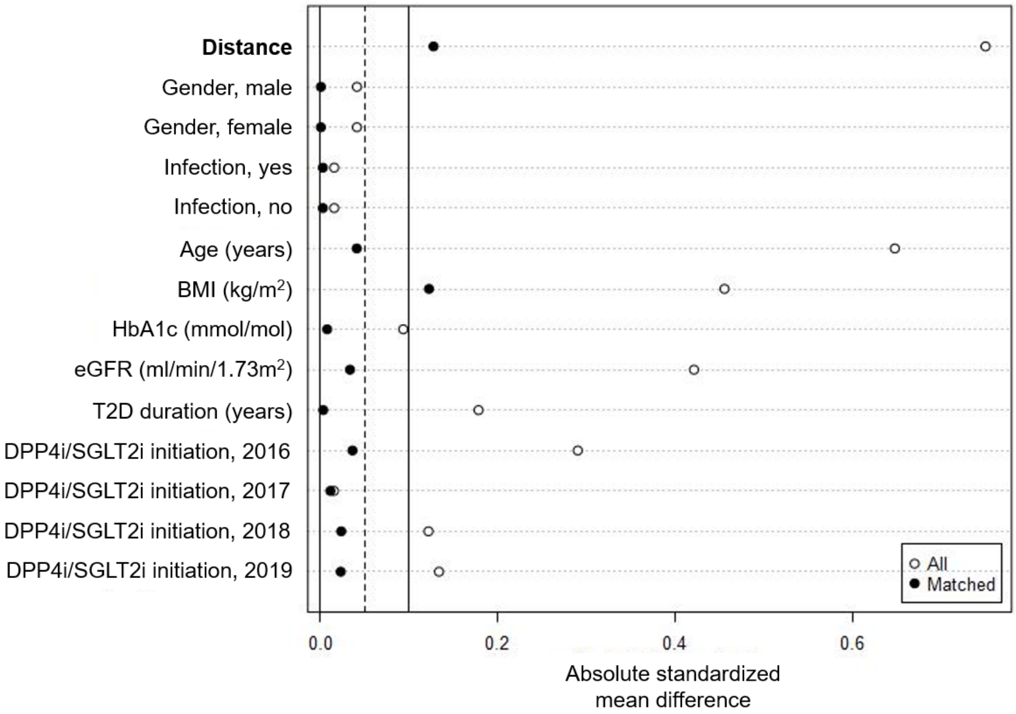


Figure 3.9: Love plot of the original and propensity score matched data.

Appendix 3.7 Pairwise correlations for all applied estimates in the application case study

Correlation plot shows the pairwise correlation of all estimates using 500 bootstrap samples as explained in Section 3.6. Estimates of the CaT and PSM as well as the estimates of the POA-IV and POA-CF are highly correlated.

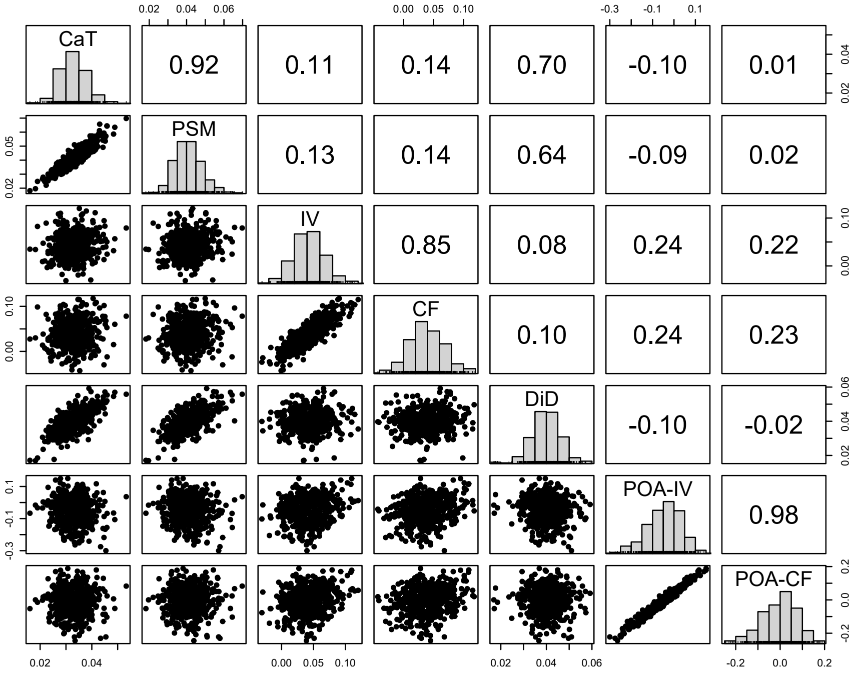
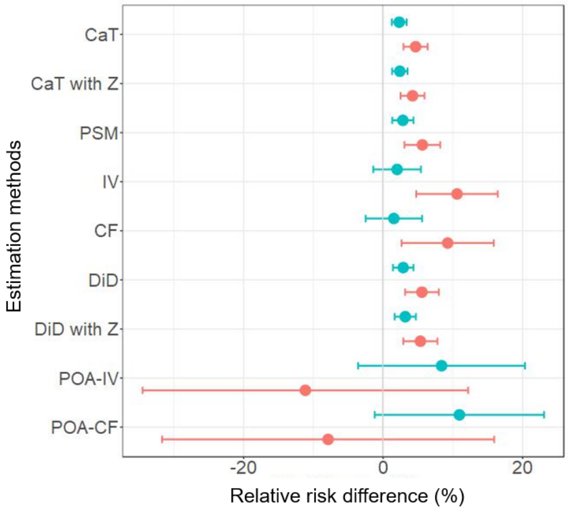


Figure 3.10: Correlation plot of all bootstrapped estimates of the application case study.

Appendix 3.8 Results of sex-stratified application case study

Additionally to the application case study outlined above, the triangulation analysis was repeated on sex-stratified cohorts. Results of the sex-stratified analysis are given in Figure 3.11. A previous study using propensity score matching and CPRD has found that female-gender is a risk factor for experiencing genital infections after the initiation of SGLT2i (versus DPP4i). [157] The results summarized in Figure 3.11 show slightly elevated but not significantly different risk for women on SGLT2i for most methods except POA-IV and POA-CF. Reasons for the differing results compared to other estimates could be that Y_0 is confounded by $\mathbf{U}$, as explained in Section 3.6. Furthermore, all IV and CF estimates show similar risk for genital infections on DPP4i and SGLT2i with the wider 95% CI encompassing 0. The estimates of CaT and PSM do not take unmeasured confounding into account, but are supported by the very similar DiD results.



Risk difference estimates in % [95% CI]

Male (n = 3578)	Female (n = 2421)
2.31 [1.24; 3.37]	4.66 [2.92; 6.39]
2.40 [1.30; 3.50]	4.22 [2.50; 5.93]
2.83 [1.31; 4.35]	5.62 [3.05; 8.19]
2.00 [-1.41; 5.42]	10.59 [4.75; 16.43]
1.54 [-2.51; 5.58]	9.26 [2.65; 15.87]
2.88 [1.42; 4.34]	5.57 [3.16; 7.97]
3.18 [1.66; 4.70]	5.34 [2.90; 7.79]
8.38 [-3.58; 20.33]	-11.15 [-34.49; 12.19]
10.93 [-1.20; 23.06]	- 7.89 [-31.69; 15.92]

Figure 3.11: Estimation results of the sex-stratified analysis.

Chapter 4

Just what the doctor ordered: An evaluation of provider preference-based Instrumental Variable methods in observational studies, with application for comparative effectiveness of type 2 diabetes therapy

Laura M. Güdemann, John M. Dennis, Beverley M. Shields & Jack Bowden
on behalf of the MASTERMIND consortium

Planned submission in: Statistics in Medicine

Pre-published at: `https://arxiv.org/abs/2308.08020`

Author contribution

LMG conceived the idea for the study and JB helped refining the research question of this paper. LMG and JB developed the proposed method for constructing preference-based instrumental variable. LMG conducted the analysis and drafted the original version of the paper which all authors helped to edit. All authors read and approved the final version of the manuscript.

4.1 Abstract

The Instrumental Variable approach provides a possibility to address bias due to unmeasured confounding when estimating treatment effects using observational data. As instrument prescription preference of individual healthcare providers has been proposed. Because prescription preference is hard to measure and often unobserved, a surrogate measure as proxy, constructed from available prescription data, is often required for Instrumental Variable analysis. Different construction methods for this proxy measure are possible, such as simple rule-based methods which make use of the observed treatment patterns of each provider, or more complex model-based methods that employ formal statistical models to explain the treatment behaviour by taking the data structure of measured confounders into consideration. The choice of construction method relies on aspects such as data availability within provider, missing data in measured confounders, and possible changes in prescription preference over time.

In this paper we conduct a comprehensive simulation study to evaluate different construction methods for proxy measures of provider prescription preference under different data conditions, including: different provider sizes, missing covariate data, and provider preferences that change over time. We additionally propose a novel model-based construction method that utilizes a mixed effect model with a random intercept for provider ID and a random slope for prescription time, to address between provider differences and change in prescription preference simultaneously. All presented construction methods are exemplified in a case study analysing the relative glucose lowering effect of two type 2 diabetes treatments in observational data, to showcase their different data requirements and to triangulate estimations results.

Our study shows that Instrumental Variable methods using provider preference can be a useful tool for causal inference from observational health data. The choice of construction method should be driven by the data condition at hand. Our proposed method is capable of estimating the causal treatment effect without bias in case of sufficient prescription data per provider, changing prescription preference over time and non-ignorable missingness in measured confounders.

4.2 Introduction

When comparing the relative effectiveness of treatments in observational data, the Instrumental Variables method (IV) provides a possibility to address bias in the estimation of treatment effects due to unmeasured confounding. The approach aims to create pseudo-randomized treatment assignment based a suitable instrument for which three main assumptions are necessary. The chosen instrument must strongly predict the treatment decision, be independent of unmeasured confounders and should only affect the outcome through influencing the treatment decision. Korn and Baumrind [80] proposed to utilize healthcare provider prescription preference (PP) as an IV. This instrument has been widely applied in health research areas such as in comparative effectiveness studies for cancer, cardiovascular diseases and mental health. [59]

Preference-based IVs have been constructed at three different healthcare provider levels in the literature: regional [83], hospitals, practices (or other institutions) [86], or at the individual physician level [88]. In order for PP to be a valid instrument, prescription habits must differ across providers in a manner that cannot be purely explained by patient characteristics which are prognostic for the studied disease, or regional variation in treatment guidelines. Provider preference must also be unrelated to the use of other medical interventions that might influence the outcome. Lastly, it is necessary that patients are assigned to provider independently of their prescription pattern. [80, 92]

In most routine clinical databases, the reason underlying a treatment decision is not systematically recorded, meaning the true prescription preference of a healthcare provider for one treatment over another is unknown. It is also likely to be a non-binary and evolving variable, representing the strength of a provider's belief in what constitutes the best treatment option for a patient at a particular point in time. Surveys have been designed in order to elicit PP information [67, 90] but it remains a difficult quantity to accurately and unbiasedly measure. [89] Therefore, comparative effectiveness studies based on observational data typically do not incorporate data on true provider preference of one drug over the other. In most IV studies using preference-based instruments, PP is instead substituted with a proxy variable that is somehow estimated from the data. This is referred to as the 'proxy design' [58, 59, 91] which is represented in Figure 4.1. Here, PP is depicted as valid but unmeasured instrument. It is approximated with the variable Z which utilizes provider's manifest and observed prescribing behaviour to reflect on PP. The outcome variable, measured and unmeasured confounders are denoted with Y, $\mathbf{W}$ and $\mathbf{U}$ respectively.

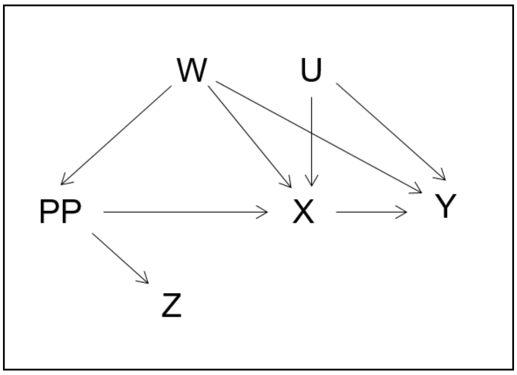


Figure 4.1: Causal diagram of the proxy design. The true underlying provider preference is not measured but instead approximated with the proxy variable Z*.*

Different methods to construct the instrument have been proposed in the literature. These can be differentiated into two general groups: simple rule-based methods that utilize subsets of the observed treatment decision data to derive Z at the point a given patient is treated; and more complex methods derived from fitting formal statistical models to the full data (encompassing treatment decisions, outcomes and measured confounders). When multiple proxy construction methods for Z are available, Brookhart et al. [92] suggest to chose the one that appears to be most strongly related to the observed treatment decision, among those that are unrelated to measured confounders. [92] This is motivated by the desire to minimise the measurement error of Z measuring PP and makes the consideration of additional aspects about prescription preferences necessary.

The prescribing preferences of providers undoubtedly contain a dynamic aspect due to the accumulating personal positive or negative experience with administering a drug over time, as well as external factors such as treatment guidelines from health authorities, marketing activities of pharmaceutical companies, and efficacy and safety information from clinical trials. [92, 93, 94, 95, 96] Abrahamowicz et al. [93] propose a construction method for Z that aims to identify if a provider changes preference and at what point in time the change took place. If such a change point is identified, Z is constructed separately for patients treated before and after the change. [93] Clearly, when implementing a particular model-based method, it is necessary to ensure that sufficient data per provider is available to support it. [92]

Valid IV estimates generally also depend on including treatment effect confounders that are associated with the instrument in the analysis. Missing data are common in observational studies and often dealt with by complete case analysis or applying imputation methods. These strategies can lead to bias in the IV estimates, in case of non-ignorable missingness for which the missingness depends on the unmeasured value itself or other unobserved variables. [45, 97, 230] In this challenging setting, Ertefaie et al. [97] propose a model-based method for constructing Z and show using theoretical arguments and a simulation study that the method is capable of producing unbiased treatment effect estimates.

The aim of this study is a state of the art evaluation of the performance of different construction methods of preference-based instruments with respect to these three data structure aspects and a focus on the more complex model-based approaches. Additionally, we propose an extension of the method by Ertefaie et al. [97] to accommodate non-ignorable missingness as well as possible change in prescription preference. In Section 4.3 methods for constructing the proxy variable Z for PP are described, with a focus on the model-based approaches. In Section 4.4, the performance of these methods is evaluated in a simulation study which allows for a change in prescription preference, different data availability and different missingness mechanisms for measured confounders. In Section 4.5 all construction methods of Z are applied to primary care data from the Clinical Practice Research Datalink (CPRD) for a comparative effectiveness study comparing two oral type 2 diabetes (T2D) agents - Sodium-glucose Cotransporter-2 Inhibitors (SGLT2i) versus Dipeptidyl peptidase-4 Inhibitors (DPP4i) - on their ability to lower blood glucose levels (HbA1c mmol/mol). Lastly, Section 4.6 concludes the main points of the simulation and application case study and highlights limitations as well as further research possibilities.

4.3 Constructing preference-based Instrumental Variables

In this section we give an overview of the possibilities for constructing a provider preference proxy variable seen in the literature. We categorize them into two groups based on their use of data on treatment behaviour and measured confounders. Simple rule-based approaches make use of the observed treatment patterns of each provider, while more complex model-based methods use formal statistical models to explain the treatment behaviour, additionally taking the data structure of measured confounders into consideration. The latter category of construction methods focuses on important aspects of preference-based IVs such as possible change in PP over time and the existence of non-ignorable missingness in confounder data. We will focus predominantly on the latter model-based case in the simulation study and application case study of Section 4.4.

4.3.1 Notation

We assume a study population of N patients, who are clustered into $j = 1, \ldots, J$ disjoint sets representing distinct treatment decisions. Provider j treats $i = 1, \ldots, n_j$ patients, so that $N = \sum_{j=1}^{J} n_j$. Within each provider, the patients' index i is assumed to coincide with the order in which they have been treated, from first to most recent. The outcome of interest for patient i of provider j is denoted by Y_{ji}. Likewise, binary treatment variable X_{ji} denotes whether a patient receives treatment A ($X_{ji} = 0$) or treatment B ($X_{ji} = 1$). Confounders are classified as either measured or unmeasured, and are represented by the G- and M-length vectors $\mathbf{W_{ji}}$ = $(W_{1ji}, \ldots, W_{Gji})$ and $\mathbf{U_{ji}}$ = $(U_{1ji}, \ldots, U_{Mji})$, respectively. Let the variable PP_{ji} represent the *true* underling preference for treatment B over treatment A of provider j at the point they treat patient i. We assume that provider preference satisfies the instrumental variable assumptions, either marginally or conditional on $\mathbf{W_{ji}}$. Finally, let Z_{ji} represent a data-derived *estimate* for PP_{ji} which will be used as a proxy measure. In the following sections, the indexes i and j will be omitted from the notation if explanations are not specific to certain providers or individuals.

A visual summary of how each method constructs the IV Z_{ji} within a hypothetical provider j is provided in Figure 4.2 for the rule-based methods and in Figure 4.3 for the established model-based approaches. Their respective x-axes denote the $i = 1, \ldots, n_j$ patients treated by provider j in sequential order. Prescription decisions X_{ji} are given on the y-axes and indicated with a x symbol. The true underlying provider preference PP_{ji} is indicated by the solid line. The corresponding derived instrument is marked with a circle. Shaded areas indicate which treatment data is utilised in the derivation of the respective Z_{ji}. It also clarifies for which patients Z_{ji} cannot be calculated.

Constructing Z is a means to an end, the end being to perform an IV analysis to overcome confounding between observed values of X and Y. Throughout this paper we will implement this analysis within the classic Two-Stage Least Squares (TSLS) framework by doing the following:

1. Use data $\mathbf{W}$ and X to derive instrument Z;
2. Regress: X on Z and $\mathbf{W}$ to obtain a predicted value $\hat{X}$;
3. Regress: Y on $\hat{X}$ and $\mathbf{W}$ to obtain an estimate for the causal effect of receiving treatment A (X = 0) versus B ($X = 1$) on the outcome

The focus of this section is on different methods for performing step 1.

4.3.2 Rule-based approaches

Brookhart et al. [92] proposed to conceptualize Z_{ji} with the most recent prescription. That is, for each patient i and treated by provider j, $Z_{ji} = X_{ji-1}$. Therefore, Z_{ji} is a binary variable and is calculable for all patients treated by provider j *except* the first one. As Z_{ji} is constructed using the most recent prior treatment decision, it can reflect on true changes in provider preference but also random fluctuations from patient to patient not indicative of a genuine change or trend. [92, 93, 100] This method will be referred to as *IV prevpatient* and is visualized in panel A of Figure 4.2.

A generalisation of this construction method calculates Z_{ji} as the proportion of patients who received a given treatment, say treatment B, among the previous b treated patients, so that

$$\mathsf{Z}_{ji} = \frac{1}{b} \sum_{d=i-(b-1)}^{i-1} \mathsf{X}_{jd} \quad = \quad b\bar{\mathsf{X}}_i(b).$$

This instrument can take on values in the range of 0% and 100% as opposed to being a binary variable. [98, 100, 231] We refer to this method as *IV prevbpatient* by setting b in the following sections equal to 2, 5, and 10. This method can theoretically reflect on changes in the providers preference depending on how close to the time of preference change patient i is treated, but is based on more data. Clearly Z_{ji} cannot be calculated for the first b patients within each provider, which means that some data is lost. A calculation example for $b = 5$ is given in panel B of Figure 4.2.

A closely related variation of this method is to construct Z_{ji} using *all* previous prescriptions for treatment B, so that:

$$\mathsf{Z}_{ji} = \frac{1}{i-1} \sum_{d=1}^{i-1} \mathsf{X}_{jd} \quad = \quad (i-1)\bar{\mathsf{X}}_i(i-1).$$

It can be calculated for all patients (except the first) and will be referred to as *IV allprevprop*. [92] This method is summarized in panel C of Figure 4.2.

For the previous methods Z_{ji} is calculated individually for each patient i within provider j. Alternatively, all prescription data can be utilized to derive a single instrument for the provider. This leads to

$$\mathsf{Z}_{ji} = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathsf{X}_{ji} \quad = \quad n_j\bar{\mathsf{X}}_j = \mathsf{Z}_j,$$

which can be calculated for $i = 1, \dots, n_j$ and lies between 0% and 100%. This will be referred to as *IV allprop*. It is possible to dichotomize Z_{ji} and created a binary instrument with the median [231] or mean empirical value of all practitioners, so that in the case of the median:

$$\mathsf{Z}_{ji} = \begin{cases} 0, & \text{if } n_j\bar{X}_j \leq \text{Median}(n_1\bar{X}_1, \dots, n_J\bar{X}_J) \\ 1, & \text{otherwise.} \end{cases}$$

These methods will be referred to as, *IV alldichmedian* and *IV alldichmean* respectively. A minimum provider size $n_{j,\min} = 2$ is needed to apply this method and its dichotomized versions. The methods are represented in panel D of Figure 4.2.

For the estimation of the treatment effect with any given preference-based IV, the data of patients for which Z_{ji} cannot be calculated are excluded. This has clear implications for the efficiency of each method.

4.3.3 Model-based approaches

We now introduce two model-based approaches to construct Z: the method of Ertefaie et al. [97] (henceforth the 'Ertefaie' method) which fits a multi-level model and is capable of dealing with non-ignorable missingness in a measured confounder; and secondly, the method of Abrahamowicz et al. [93] (henceforth the 'Abrahamowicz' method) which tests for and then potentially models a change point in provider preference over time. We additionally propose an extension of the Ertefaie method that allows for time trend in provider preference using a mixed model with a random intercept and slope.

The Ertefaie method

Ertefaie et al. [97] propose a procedure for the estimation of a valid treatment effect utilizing IVs based on provider preference in the presence of baseline characteristics with non-ignorable missingness. For this approach the set of measured baseline characteristics is subdivided in $\mathbf{W_{obs,ji}}$ for all confounders fully observed, and $\mathbf{W_{miss,ji}}$ denoting all confounders which are not completely recorded for all i. In the following, $\mathsf{G}_{\mathsf{miss}}$ and $\mathsf{G}_{\mathsf{obs}}$ denotes the number of confounders in $\mathbf{W_{miss,ji}}$ and $\mathbf{W_{obs,ji}}$ respectively.

The instrument Z_{ji} is estimated from a generalized random multilevel model, regressing X_{ji} on $\mathbf{W_{obs,ji}}$ and $\mathbf{W_{miss,ji}}$. The model includes the random intercept γ_{0j} for each provider (provID) and is estimated using a complete case dataset on all measured confounders:

$$\begin{aligned}\mathsf{Logit}(\mathsf{Pr}[\mathsf{X}_{\mathsf{ji}} = 1|\mathbf{W_{obs,ji}}, \mathbf{W_{miss,ji}}, \mathsf{provID}_{\mathsf{ji}}]) = \gamma_0 + \gamma_{\mathsf{0j}} + \gamma_{\mathbf{W_{miss,ji}}}\mathbf{W_{miss,ji}} +& \\ \gamma_{\mathbf{W_{obs}}}\mathbf{W_{obs,ji}} + \varepsilon_{\mathsf{ji}}&\end{aligned} \tag{4.1}$$

where $\gamma_{\mathbf{W_{miss,ji}}}$ and $\gamma_{\mathbf{W_{obs,ji}}}$ are column vectors of sizes $\mathsf{G}_{\mathsf{miss}} \times 1$ and $\mathsf{G}_{\mathsf{obs}} \times 1$, and $\mathbf{W_{miss,ji}}$, $\mathbf{W_{obs,ji}}$ are matrices of sizes $\mathsf{N} \times \mathsf{G}_{\mathsf{miss}}$ and $\mathsf{N} \times \mathsf{G}_{\mathsf{obs}}$. Ertefaie et al. [97]

propose to estimate the instrument for provider j based on the relative position of random intercept $\hat{\gamma}_{0j}$ in the entire distribution, so that:

$$Z_{ji} = Z_j = \begin{cases} 1, & \text{if}\,(\text{expit}(\hat{\gamma}_{0j}) > \text{expit}(\text{Median}\,\{\hat{\gamma}_{01}, \ldots, \hat{\gamma}_{0J}\})) \\ 0, & \text{otherwise.} \end{cases}$$

The visualisation of this approach in Figure 4.3 (top) shows that each patient within provider j will receive the same value for Z_{ji} and that the instrument can be estimated for all i in j. The minimum sample size for this construction method can not be precisely determined from data availability as for the rule-based methods, but depends on the estimation of the mixed effect model. An often cited rule of thumb for minimum sample size is Kreft's 30/30 rule which requires data on at least 30 provider and 30 patients within each provider. [232] Further discussion on this can be found for example in Snijders and Bosker [233] or Hox and van de Schoot [234]. In Section 4.4 the effects of different provider sizes on the estimation performance will be discussed and we employ this construction method as well as its extension method if a provider treats at least two patients.

An advantage of the Ertefaie method is that, despite the IV derivation being based on complete case data, the final analysis and TSLS estimation of the treatment effect is employed on the full data, only using $\mathbf{W}_{\mathbf{obs}}$ in the first and second stage models. The only proviso is that a random intercept and hence Z_{ji} can be derived for a given provider. If in fact there are no complete case individuals in a given provider, the provider must be excluded from the analysis.

The Ertefaie method relies on three assumptions. The first assumptions requires missingness to occur at provider level unrelated to unmeasured confounders and the treatment and at individual patient level, which can be dependent on measured and unmeasured confounders. This assumption is plausible for studies in which missingness varies across providers for example due to staff or management policies. In a sensitivity analysis the authors show that the approach is still valid under moderate violations of this assumption. [97] The second assumption states that the effect of $\mathbf{U}$ on X should not vary by provider. As this is not testable with the data at hand, the authors suggest that it is possible to make an educated guess by checking whether measured confounders violate this assumption. This could be checked with a generalized mixed-effect model to predict the treatment decision including measured confounders and provID as random intercept and random slope. Significant random slopes would indicate that this assumption might be violated. The third assumption (positivity) indicates that each provider has a non-zero probability in theory of seeing patients with any given observed characteristics. The authors verify this approach in a simulation study with non-ignorable missingness in one measured confounder. The treatment effect estimation results are unbiased even in case of non-ignorable missingness and have lower standard errors compared to standard IV approaches using complete case or multiple imputed datasets. A sensitivity analysis confirms that the approach can even deal with 'high rates' of non-ignorable missingness. [97] Since the method makes use of a complex mixed effect model and the prescription data over all providers simultaneously, it is hard to relate the derivation of Z to the observed treatment data X in any given provider. Nevertheless, the method is summarized in panel A of Figure 4.3. When summarizing results in Section 4.4 and Section 4.5 it will be referred to as *IV ePP*.

The Abrahamowicz method

In order to address concerns about variance inflation when using only one previous prescription to construct Z, Abrahamowicz et al. [93] propose a more complex modification of IV prevpatient that aims to detect and account for a change in provider preference over time, yet uses more prescription data to achieve smaller estimation variance. The procedure is applied for each of the J providers individually in the following 4 step procedure. We follow the previous convention by assuming patients within provider j are ranked by calendar time from 1 to n_j.

Step 1: To test if a provider changes preference, a reference no-change model is estimated from a multivariable logistic regression of X_{ji} with $g = 1, \dots, G$ measured covariates.

$$\mathrm{Logit}(\Pr[X_{ji} = 1]) = \beta_0 + \sum_{g=1}^{G} \beta_g W_{gji}.$$

From this model the deviance $D(0)_j$ for provider j is extracted.

Step 2: This step is applied to test for a change in preference of provider j and is further split up into three iterations.

Step 2.1: The difference in prescription proportions for each patient $\mathsf{i} = 3, \ldots, \mathsf{n_j} - 3$ is calculated by subtracting the proportion of $\mathsf{X_{ji}} = 1$ of the 'later patients' with $\mathsf{k} > \mathsf{i}$ from the proportion of 'earlier patients' with $\mathsf{k} \leq \mathsf{i}$:

$$\mathsf{d_{ji}} = \underbrace{\frac{\sum_{1\leq k\leq i}}{i}}_{\text{earlier patients}} - \underbrace{\frac{\sum_{i+1\leq k\leq n_j}}{n_j - i}}_{\text{later patients}}.$$

As this difference in proportions is calculated from the third to the third last patient treated by provider j, the minimum number of patients needed within each provider is 5. The provider preference is assumed to be constant across the observed time period if $|\mathsf{d_{ji}} < 0.2|\ \forall \mathsf{i}$ in $\mathsf{i} = 3, \ldots, \mathsf{n_j} - 3$. For all other provider with at least one $|\mathsf{d_{ji}} \geq 0.2|$ further investigations are conducted to identify possible change in preference. In the next steps the time of the preference change is identified.

Step 2.2: A change-time model is estimated for each patient i with $|\mathsf{d_{ji}} \geq 0.2|$:

$$\mathsf{Logit(Pr[X_{ji} = 1])} = \beta_0 + \sum_{\mathsf{g}=1}^{\mathsf{G}} \beta_{\mathsf{g}} \mathsf{W_{gji}} + \eta \mathbb{1}(\mathsf{k} > \mathsf{i}).$$

In addition to the no-change model, this model includes a binary variable indicating patients prescribed after i with $\mathbb{1}(\mathsf{k} > \mathsf{i}) = 0$ if $\mathsf{k} \leq \mathsf{i}$ and $\mathbb{1}(\mathsf{k} > \mathsf{i}) = 1$ if $\mathsf{k} > \mathsf{i}$. Note that for the most recently prescribed patient $\mathsf{i} = \mathsf{n_j}$, $\mathbb{1}(\mathsf{k} > \mathsf{i}) = 0$. The approach will therefore not be able to test for a change in preference directly after the last prescribed patient of provider j. The parameter η is the average adjusted difference in the propensity of provider j to prescribe $\mathsf{X_{ji}} = 1$ between 'earlier patients' and 'later patients'. From the change-time models the deviances $\mathsf{D_j(i)}$ are extracted.

Step 2.3: In order to identify after which prescription a possible change in preference took place, the optimal change-time model and the time of change $\mathsf{i}^\star$ is defined as

$$\mathsf{min(D_j(i) = D_j(i^\star))}.$$

Step 3: In this step the fits of the no-change model and the optimal change-time model are compared using the Akaike information criterion, which is calculated as $\mathsf{AIC} =$ deviance $+2\mathsf{S}$, S being the number of parameters estimated in the model. For the change-time model two additional parameter are estimated, η and the optimal change time $i^\star$. Therefore, the no-change model is only considered to have a better fit if

$$D(0)_j < D_j(i^\star) + 4. \tag{4.2}$$

Step 4: Providers are identified to *not* have a change in preference if, as explained in step 2.1 $|d_{ji} \geq 0.2| \; \forall i$ in j or if the no-change model is the best fit explained in step 3. In this case Z_{ji} for all patients in provider j are constructed as IV allprevprop explained in Section 4.3.2. Providers *are* identified to have a change in preference if the conditions in step 2.1 and step 2.3 are satisfied. It is then assumed that the change in preference takes place between $i^\star$ and $i^\star + 1$. In this case Z_{ji} is constructed by firstly subdividing the prescription data into two subgroups, before and after the change time. The two groups will encompass patients $i = 1, \ldots i^\star$ and patients $i = i^\star + 1, \ldots n_j$ respectively. Following this, Z_{ji} is calculated as IV allprevprop in both subgroups limited to the treatment information from those subgroups.

Panel B of Figure 4.3 shows a calculation example for Z_{ji}, in two cases where $i^\star$ is correctly and incorrectly identified. The calculation of Z_{ji} for all patients after the change will depend on the identification of $i^\star$. Additionally, the graph points out that for providers who have a change identified by the method, Z_{ji} cannot be calculated for the two patients $i = 1$ and $i = i^\star + 1$. Lastly, it should be noted that the identification of a change in preference relies on the assumption that provider will only change their preference once within the observed time period. When summarizing results in consecutive sections the method will be referred to as *IV star*.

Extension of the Ertefaie method

We propose an extension to the Ertefaie method, *IV ePP (rirs)*, that utilizes a random intercept random slope model to construct Z. This approach aims to combine the original strategy for estimating unbiased treatment effects even under non-ignorable missingness for measured confounders and the principle idea underpinning the Abrahamowicz method to account for the possibility of changing prescription preference over the observed period, albeit implemented in a more straightforward model. Specifically, the instrument Z_{ji} is constructed from a generalized random intercept, random slope model for the treatment decision:

$$\begin{aligned}\mathsf{Logit}(\mathsf{Pr}[\mathsf{X}_{\mathsf{ji}} = 1|\mathbf{W}_{\mathbf{obs,ji}}, \mathbf{W}_{\mathbf{miss,ji}}, \mathsf{provID}_{\mathsf{ji}}, \mathsf{T}_{\mathsf{ji}}]) = \gamma_0 + \gamma_{0\mathsf{j}} + (\gamma_{\mathsf{Tj}} + \gamma_{\mathsf{T}})\mathsf{T}_{\mathsf{ji}} + \\ \gamma_{\mathbf{W}_{\mathbf{miss}}}\mathbf{W}_{\mathbf{miss,ji}} + \gamma_{\mathbf{W}_{\mathbf{obs}}}\mathbf{W}_{\mathbf{obs,ji}} + \varepsilon_{\mathsf{ji}},\end{aligned} \tag{4.3}$$

where T_{ji} represents a variable indicating the time of prescription and $\gamma_{\mathbf{W}_{\mathbf{miss}}}$, $\gamma_{\mathbf{W}_{\mathbf{obs}}}$, $\mathbf{W}_{\mathbf{obs,ji}}$, $\mathbf{W}_{\mathbf{miss,ji}}$ and $\mathsf{provID}_{\mathsf{ji}}$ are defined as before. From this model the following fitted values $\hat{\Theta}_{\mathsf{ji}}$ are derived with the estimated global intercept $\hat{\gamma}_0$, random intercept $\hat{\gamma}_{0\mathsf{j}}$, the global slope for T_{ji} $\hat{\gamma}_{\mathsf{T}}$ and the respective random slope $\hat{\gamma}_{\mathsf{Tj}}$:

$$\hat{\Theta}_{\mathsf{ji}} = \hat{\gamma}_0 + \hat{\gamma}_{0\mathsf{j}} + (\hat{\gamma}_{\mathsf{T}} + \hat{\gamma}_{\mathsf{Tj}})\mathsf{T}_{\mathsf{ji}}. \tag{4.4}$$

Finally, the instrument is constructed by applying the following binary transformation to $\hat{\Theta}_{ji}$:

$$\mathsf{Z}_{\mathsf{ji}} = \begin{cases} 1, & \text{if } \left(\mathsf{expit}(\hat{\Theta}_{\mathsf{ji}}) > \mathsf{expit}(\text{Median}\left\{\hat{\Theta}_{11}, \ldots, \hat{\Theta}_{\mathsf{Jn_J}}, \right\})\right) \\ 0, & \text{otherwise.} \end{cases}$$

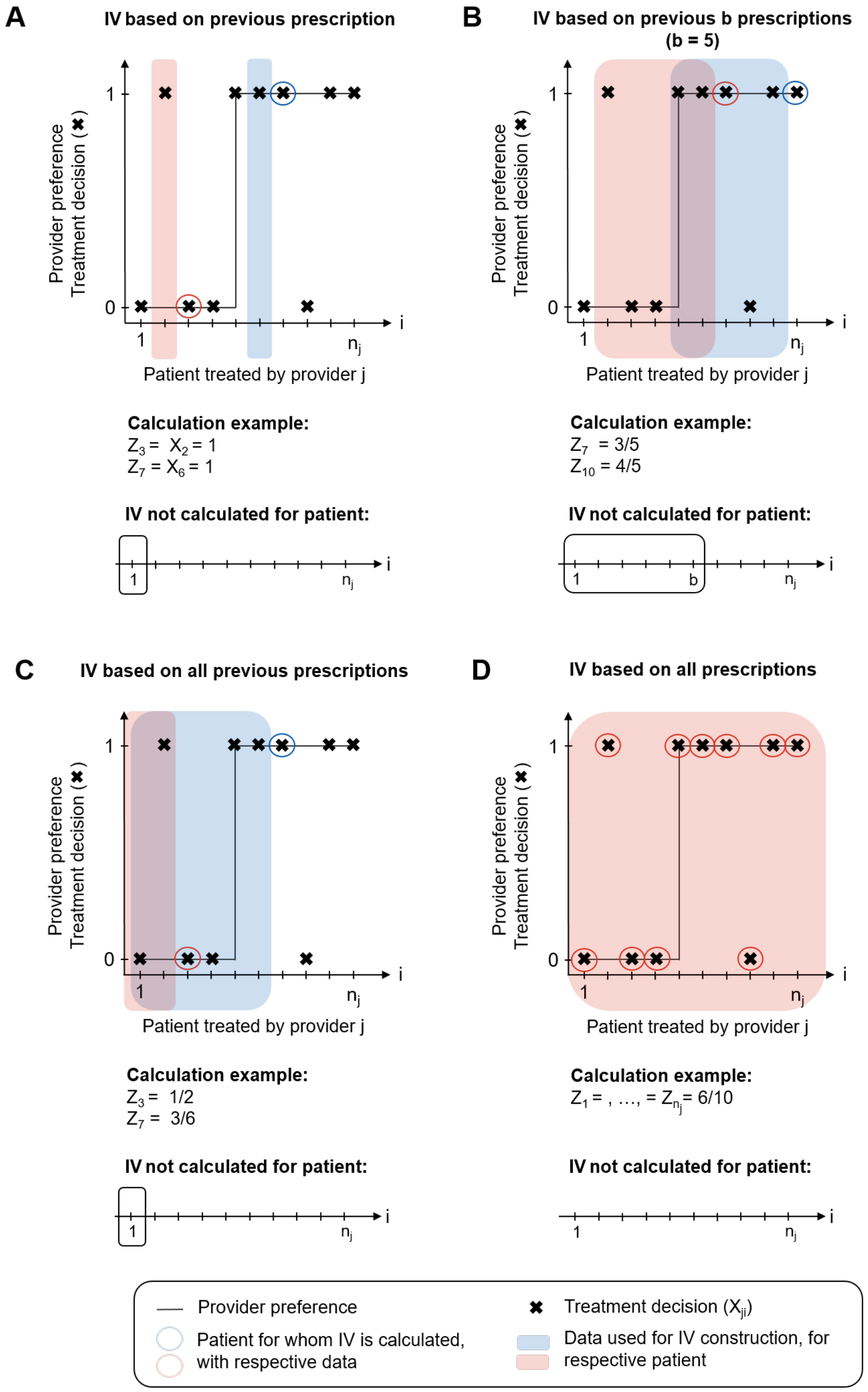

A
IV based on previous prescription
Provider preference
Treatment decision (✖)
1
0
1
n_j
i
Patient treated by provider j
Calculation example:
$Z_3 = X_2 = 1$
$Z_7 = X_6 = 1$
IV not calculated for patient:
1
n_j
i
B
IV based on previous b prescriptions (b = 5)
Provider preference
Treatment decision (✖)
1
0
1
n_j
i
Patient treated by provider j
Calculation example:
$Z_7 = 3/5$
$Z_{10} = 4/5$
IV not calculated for patient:
1
b
n_j
i
C
IV based on all previous prescriptions
Provider preference
Treatment decision (✖)
1
0
1
n_j
i
Patient treated by provider j
Calculation example:
$Z_3 = 1/2$
$Z_7 = 3/6$
IV not calculated for patient:
1
n_j
i
D
IV based on all prescriptions
Provider preference
Treatment decision (✖)
1
0
1
n_j
i
Patient treated by provider j
Calculation example:
$Z_1 = , \ldots, = Z_{n_j} = 6/10$
IV not calculated for patient:
1
n_j
i
Provider preference
Patient for whom IV is calculated, with respective data
✖ Treatment decision (X_{ji})
Data used for IV construction, for respective patient

Figure 4.2: Visualization of the rule-based preference-based IV construction methods with corresponding calculation example. Abbreviations used for the methods are A: IV prevpatient, B: IV prevbpatient, C: IV allprevprop, D: IV all-prop, IV alldichmean, IV alldichmedian.

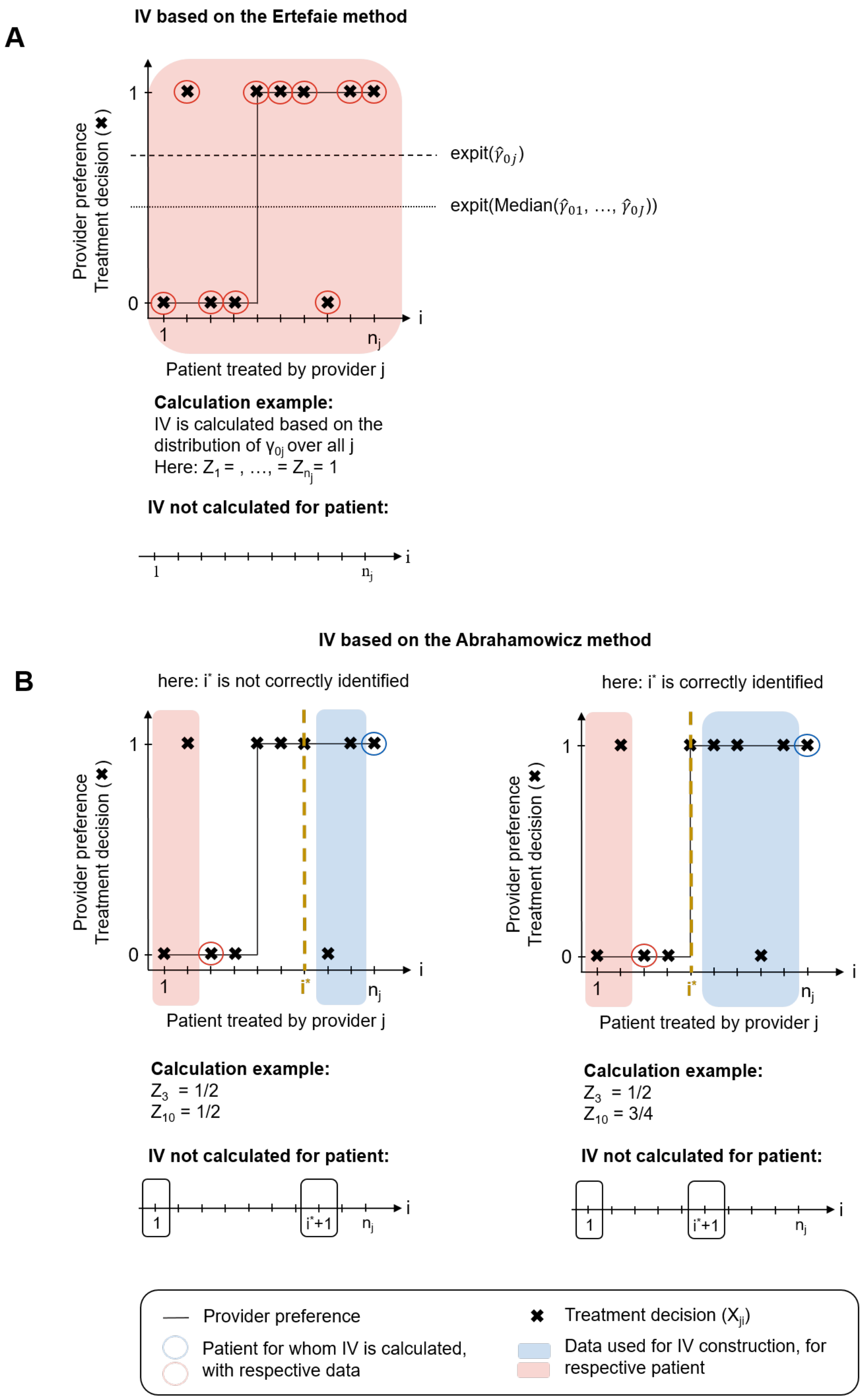

A
IV based on the Ertefaie method
Provider preference
Treatment decision (✖)
1
0
expit($\hat{\gamma}_{0j}$)
expit(Median($\hat{\gamma}_{01}$, ..., $\hat{\gamma}_{0J}$))
1
n_j
i
Patient treated by provider j
Calculation example:
IV is calculated based on the distribution of γ_{0j} over all j
Here: Z_1 = , ..., = Z_{n_j} = 1
IV not calculated for patient:
1
n_j
i
IV based on the Abrahamowicz method
B
here: i^* is not correctly identified
here: i^* is correctly identified
Provider preference
Treatment decision (✖)
1
0
1
i^*
n_j
i
Patient treated by provider j
Calculation example:
Z_3 = 1/2
Z_{10} = 1/2
IV not calculated for patient:
1
i^*+1
n_j
i
Calculation example:
Z_3 = 1/2
Z_{10} = 3/4
IV not calculated for patient:
1
i^*+1
n_j
i
Provider preference
Patient for whom IV is calculated, with respective data
✖ Treatment decision (X_{ji})
Data used for IV construction, for respective patient

Figure 4.3: Visualization of the established model-based preference-based IV construction methods. The Abrahamowicz method is shown with correctly identified (left side) and incorrectly identified (right side) change time i⋆. Abbreviations used for the methods are A: IV ePP, B: IV star.

4.4 Simulation study

Several simulation studies investigating different aspects of provider preference-based IVs can be found in the literature. Ionescu-Ittu et al. [82] assesses the performance of using prescription data of previous patients under varying instrument strengths. Abrahamowicz et al. [93] and Ertefaie et al. [97] also include simulations studies showcasing the key feature of their proposed methods, namely modelling a change in prescription preference and accounting for non-ignorable missingness.

Our simulation study aims to compare all rule-based and model-based methods for PP IV construction, but with a specific focus on the two model-based approaches proposed by Abrahamowicz et al. [93] and Ertefaie et al. [97], as well as our extension method. Results of the rule-based method are referred to and are discussed in more detail in Appendix 4.3.

4.4.1 Data generation

Population data is generated for $J = 100$ providers who treat each $1, \dots, n_j$ patients in ascending order. The two measured covariates generated for the simulation are $W_{1,ji} \sim N(\mu_{W_1}, 2)$ and $W_{2,ji} \sim N(\mu_{W_2}, 2)$ with μ_{W_1} and $\mu_{W_2} \sim N(0, 0.5)$ and additionally, one unmeasured confounder is generated with $U_{ji} \sim N(0, 1)$. The outcome of interest, Y, is simulated as a continuous variable with

$$Y_{ji} = \gamma_{Y,0} + \beta X_{ji} + \gamma_{Y,W_1} W_{1,ji} + \gamma_{Y,W_2} W_{2,ji} + \gamma_{Y,U} U_{ji} + \varepsilon_{Y,ji} \tag{4.5}$$

and the true (or causal) treatment effect is fixed at $\beta = 1$.

In order to not intentionally bias simulation results in favour of either model-based method, treatment decision data X were generated under two separate processes to be congenial for the Abrahamowicz or Ertefaie methods respectively.

Generating X under the Abrahamowicz model

Following Abrahamowicz et al. [93] we simulate the prescription preference of each provider for a given patient using a binary variable PP_{ji} which indicates the preference for treatment B ($\mathsf{X_{ji}} = 1$) if $\mathsf{PP_{ji}} = 1$. Specifically, our data generating model for $\mathsf{X_{ji}}$ is:

$$\mathsf{X_{ji}} \sim \mathsf{Bern}(\gamma_{\mathsf{X,0}} + \beta_{\mathsf{PP,ji}}\mathsf{PP_{ji}} + \gamma_{\mathsf{X,U}}\mathsf{U_{ji}} + \gamma_{\mathsf{X,W_1}}\mathsf{W_{1,ji}} + \gamma_{\mathsf{X,W_2}}\mathsf{W_{2,ji}}). \tag{4.6}$$

In order to induce a degree of stochasticity into the variable for PP, we use the following procedure:

- The initial preference of provider j is simulated with $\mathsf{PP_{initial,ji}} \sim \mathsf{Bern}(0.6)$, so that 40% and 60% prefer treatment A and B respectively
- Original ‘A preferers’ change to ‘B preferers’ with a probability of 70% for some $2 \leq \mathsf{i}^* \leq \mathsf{n_j}$
- Original ‘B preferers’ change to ‘A preferers’ with a probability of 40% for some $2 \leq \mathsf{i}^* \leq \mathsf{n_j}$

The change time is initially simulated with change time $\mathsf{i}^* \sim \mathsf{U}(0.4 \times \mathsf{n_j}, 0.7 \times \mathsf{n_j})$ $\forall \mathsf{j}$ and $\beta_{\mathsf{PP,ji}} = 0.7$. This means that, on average around 57% of providers will change their preference. As in the original simulation study by Abrahamowicz et al. [93], we also simulate data under a ‘smooth’ change model. In this case $\beta_{\mathsf{PP,ji}}$ is simulated using a smooth 3-linear model as follows: the change time $\mathsf{i}^\star$ is generated as previously but the change takes place over the interval of length $\mathsf{L_j}$ where $\mathsf{L_j} \sim \mathsf{U}(0.6 \times \mathsf{n_j}, 1.2 \times \mathsf{n_j})$. For A → B switchers, $\beta_{\mathsf{PP,ji}}$ is defined as

$$\beta_{\mathsf{PP,ji}} = \mathsf{PP_{final,ji}}\left(\frac{\mathsf{i} - \mathsf{i}^\star}{\mathsf{L_j}}\right) \quad \forall \mathsf{i}^\star < \mathsf{i} < \mathsf{i}^\star + \mathsf{L_j} \tag{4.7}$$

with $\mathsf{PP_{final,ji}}$ denoting the final preference for treatment B. Also, $\beta_{\mathsf{PP,ji}} = 0 \ \forall \mathsf{i} \leq \mathsf{i}^\star$ (patients treated before change time) and $\beta_{\mathsf{PP,ji}} = \mathsf{PP_{final,ji}} \ \forall \mathsf{i} \geq \mathsf{i}^\star + \mathsf{L_j}$ (patients treated after change period has ended). For B → A switchers, $\beta_{\mathsf{PP,ji}}$ is defined as

$$\beta_{\mathsf{PP,ji}} = \mathsf{PP_{initial,ji}}\left[1 - \left(\frac{\mathsf{i} - \mathsf{i}^\star}{\mathsf{L_j}}\right)\right] \quad \forall \mathsf{i}^\star < \mathsf{i} < \mathsf{i}^\star + \mathsf{L_j} \tag{4.8}$$

with $\mathsf{PP_{initial,ji}}$ denoting the initial preference for treatment B. For patients treated before and after the change time $\beta_{\mathsf{PP,ji}} = \mathsf{PP_{initial,ji}} \ \forall \mathsf{i} \ \mathsf{i} \leq \mathsf{i}^\star$ and $\beta_{\mathsf{PP,ji}} = 0 \ \forall \mathsf{i} \ \mathsf{i} \geq \mathsf{i}^\star + \mathsf{L_j}$ holds respectively. Calculation examples for both provider types are given in Appendix 4.1.

Generating X under the extended Ertefaie model

Ertefaie et al. [97] use a mixed effect model with a random intercept to model provider preference. To compliment this we simulate treatment decision data X_{ji} using a mixed effect model analogous to Equation 4.3, including both a random intercept to model different initial preference levels and a random slope to allow for a change in PP:

$$X_{ji} \sim \text{Bern}\left(\gamma_{X,0} + \gamma_{X,0j} + (\gamma_{X,T} + \gamma_{X,Tj})T_{ji} + \gamma_{X,U}U_{ji} + \gamma_{X,W_1}W_{1,ji} + \gamma_{X,W_2}W_{2,ji}\right). \quad (4.9)$$

Here, T_{ji} increases from 1 to 12 in ascending order of i, and one could view these time points as successive calendar months. The random intercept and random slope parameters are simulated with

$$\begin{bmatrix} \gamma_{X,0j} \\ \gamma_{X,Tj} \end{bmatrix} \sim \text{N}(0, \Omega) \text{ and } \Omega = \begin{bmatrix} \sigma^2_{\gamma_{0j}} & \sigma_{\gamma_{T0j}} \\ \sigma_{\gamma_{0Tj}} & \sigma^2_{\gamma_{Tj}} \end{bmatrix}.$$

The simulation was conduced using R version 4.2.1 and the analysis of each scenario was repeated in 200 simulation runs. R code for the simulation can be found in `https://github.com/GuedemannLaura/ppIV`. Results are given for both ways of modelling the treatment decision in equation's (4.6) and (4.9).

Table 4.6 in Appendix 4.2 summarises additional information on both data generation strategies. For both strategies the variances of Y_{ji} are similar with $\text{Var}(Y_{ji}) \approx 7.7$ for the simulation generating X_{ji} under the Abrahamowicz model and $\text{Var}(Y_{ji}) \approx 6.9$ for the simulation using the extended Ertefaie model. Though, the proportion of treated patients ($X_{ji} = 1$) shows more differences with around 42% for the former and 56% for the latter simulation strategy.

4.4.2 Scenarios

Three simulation scenarios are chosen to probe different challenges for the analysis: change in provider preference over time, a variable amount of available prescription data per provider, and missing data in the measured confounders. For the first scenario, the number of treated patients per provider is chosen with $n_j = 24, 108, 408$. The second scenario involves the simulation of missing values in the measured confounder variable W_1 due to different missing mechanisms:

either missing completely at random (MCAR) or a non-ignorable missing data mechanism. The latter will be referred to as a missing not at random mechanism (MNAR). Missingness in $\mathsf{W}_{1,ji}$ is indicated with the indicator variable F_{ji} and both mechanism results in $\mathsf{p}(\mathsf{F}_{ji} = 1) \approx 40\%$.

For the MNAR mechanism the missingness is simulated using the same data generation process as Ertefaie et al. [97]. The mechanism depends on all measured and unmeasured confounders, the outcome variable and V, the provider level influence on missing data. The missingness indicator is $\mathsf{F}_{\mathsf{MNAR},ji} \sim \mathsf{Bern}(\rho_{\mathsf{F},ji})$, where

$$\begin{aligned}\rho_{F,ji} =& \frac{exp(\gamma_{F,0} + \gamma_{F,W_1} W_{1,ji} + \gamma_{F,W_2} W_{2,ji} + \gamma_{F,U} U_{ji} + \gamma_{F,Y^\star_{ji}} Y^\star_{ji})}{1 + exp(\gamma_{R,0} + \gamma_{F,W_1} W_{1,ji} + \gamma_{F,W_2} W_{2,ji} + \gamma_{F,U} U_{ji} + \gamma_{F,Y^\star_{ji}} Y^\star_{ji})} \\ & \times \frac{exp(\gamma_{F,0} + \gamma_{F,V} V_{ji} + \gamma_{F,VW_1} V_{ji} W_{1,ji} + \gamma_{F,VW_2} V_{ji} W_{2,ji})}{1 + exp(\gamma_{F,0} + \gamma_{F,V} V_{ji} + \gamma_{F,VW_1} V_{ji} W_{1,ji} + \gamma_{F,VW_2} V_{ji} W_{2,ji})}\end{aligned} \tag{4.10}$$

with $\mathsf{Y}^\star$ denoting the standardized outcome variable and $\mathsf{V}_{ji} \sim \mathsf{U}(-2, 2)$. For the third scenario, the smooth change in preference applied by Abrahamowicz et al. [93] is employed in place of the abrupt change. Table 4.1 summarizes all scenarios and clarifies how they are implemented for each data generation strategy. The focus of each scenario with a change in parameters is indicated in **bold**.

		Data generation of X	
		Abrahamowicz model	Extended Ertefaie model
Scenario 1	n_j missing data $\beta_{PP,ji}$ change in PP_{ji} type of change	**24, 108, 408** no NAs 0.7 $\forall$j some j abrupt	**24, 108, 408** no NAs
Scenario 2	n_j missing data $\beta_{PP,ji}$ change in PP_{ji} type of change	408 **no NAs, MCAR, MNAR** 0.7 $\forall$j some j abrupt	408 **no NAs, MCAR, MNAR**
Scenario 3	n_j missing data $\beta_{PP,ji}$ change in PP_{ji} type of change	408 no NAs 0.7 $\forall$j some j **abrupt and smooth**	-

Table 4.1: Summary of the simulation scenarios for the two data generation strategies of X. *The focus of each scenario and the corresponding change in parameters is highlighted in* ***bold***.

4.4.3 Results

The estimation results of the treatment effect are represented with density plots in Figure 4.4 and 4.5 for the model-based methods. Results of the rule-based methods can be found in Appendix 4.3. As useful benchmark results for the IV estimation, three additional estimates are reported

- An 'as Treated' estimate is calculated using a multivariable regression model for Y on the observed treatment decision variable X adjusted for the measured confounders only;
- IV(PP) describes the estimate using the true simulated PP as IV. Depending on the data generation process the PP variable is either explicitly simulated (Abrahamowicz model) or can be derived with the true value of Θ calculated with formula (4.4) (extended Ertefaie model) and the true simulated values for global and random intercept and slope;
- IV(PP) cc is the complete case analysis version of IV(PP).

As the latter two estimates use the true PP, their results provide useful upper bounds for the performance of any method that constructs a provider prescription preference proxy from the data. For all scenarios with NAs, the results of IV(PP) cc give additional insight in the bias caused by applying a complete case analysis. The ‘as Treated’ estimate is denoted as *observational estimate* in all result summaries. Density plots will give insights to the bias and estimation variance across different construction methods. Further results on the coverage and relative root mean squared error (RMSE) are given in Appendix 4.4 for all methods. Besides the estimation performance, Table 4.3 and 4.4 summarize the F-statistic of the first-stage regression model from the TSLS IV estimation and for the model-based construction methods. [70] For the rule-based construction methods the F-statistic tables can be found in Appendix 4.3. This information is valuable when judging the strength of the instrument as a result from the different construction methods. Often, the instrument is considered to be a weak instrument in case of F-statistic values smaller than 10. [77] Table 4.2 and 4.7 summarize the methods applied in the simulation study and their abbreviations.

Abbreviation	Method
Obs. estimate	Observational estimate, multivariable regression adjusted for measured confounders
IV(PP)	True simulated PP as IV, utilizing all data in case of missingness
IV(PP) cc	True simulated PP as IV, utilizing complete case data in case of missingness
IV ePP	IV constructed with the Ertefaie method
IV ePP (rirs)	IV constructed with our proposed extended Ertefaie method
IV star	IV constructed with the Abrahamowicz method

Table 4.2: Summary of the model-based IV construction methods and benchmarking methods applied in the simulation and their abbreviations.

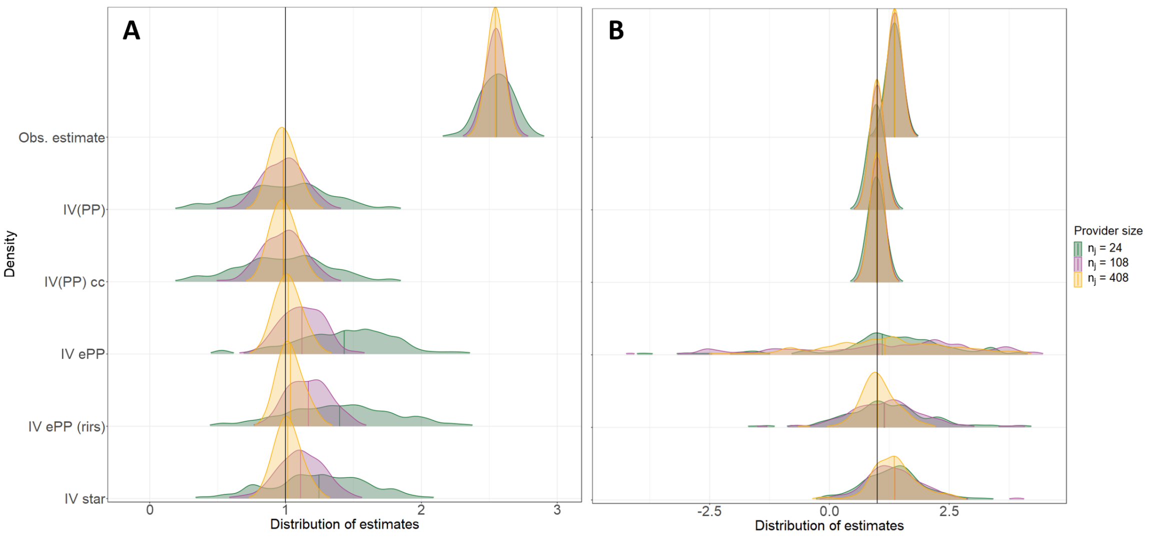


Figure 4.4: Estimation results of scenario 1: change in provider size. Panel A: estimation results for the data generation process using the Abrahamowicz model. Panel B: estimation results for the data generation process using the extended Ertefaie model to simulate change in preference. Results are summarized for the model-based construction methods of Z*.*

With scenario 1 the first aspect, different amount of available data, is investigated. The provider sizes are chosen with $\mathsf{n_j} = 24, 108, 408$ and estimation results of this scenario are summarized in Figure 4.4. Panel A shows the estimation results of the model-based construction methods for the simulation strategy of X simulating PP. Estimation of the treatment effect is more efficient with larger $\mathsf{n_j}$ and biased for smaller provider sizes. This indicates that we need sufficient data for the more complex models to adequately recover the true PP. For the Ertefaie method and its extension method, this results could be explained with the application of mixed effect model. A discussion on sample size requirements for mixed effect models can be found for example in Snijders and Bosker [233] or Hox and van de Schoot [234]. As IV(PP) is unbiased for all $\mathsf{n_j}$, the treatment effect estimates are very likely biased due to measurement error when constructing Z as proxy for PP. In Appendix 4.5 a small simulation study to evaluate the classification performance of the Abrahamowicz method IV star under different provider sizes is outlined. The rate with which the algorithm identifies provider with a change in PP correctly increases between provider sizes $\mathsf{n_j} = 10$ and $\mathsf{n_j} = 100$. This result underlines improved estimation performance of IV star for $\mathsf{n_j} = 408$. F-statistic results summarized in Table 4.3 show that all methods lead to strong instruments. Additionally, Table 4.10 in Appendix 4.4 shows that all model-based methods show good coverage in case of sufficient provider size (i.e. $n_j = 108$ and $n_j = 408$).

From Panel B of Figure 4.4 it is noticeable that the treatment effect is generally estimated with higher estimation variance when the treatment decision is simulated with a mixed effect model. Only IV ePP (rirs) estimates the treatment effect without bias, given sufficient provider size. IV ePP exhibits only small bias, but also largest estimation variance, as the model of the first step uses only a random intercept to model X. This results underlines the importance of specifying the model for PP correctly. All F-statistic results show that the constructed instrument are strong with values larger than 10.

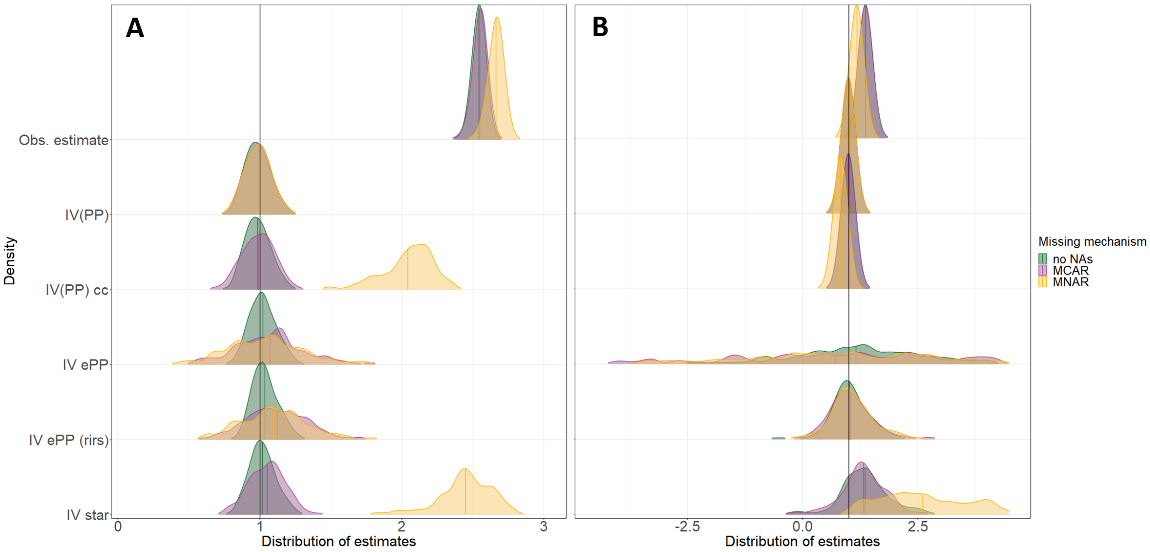


Figure 4.5: Estimation results of scenario 2: missing data mechanisms. Panel A: estimation results for the data generation process using the Abrahamowicz model. Panel B: estimation results for the data generation process using the extended Ertefaie model to simulate change in preference. Results are summarized for the model-based construction methods of Z.

With scenario 2 the effect on the estimation results due to different mechanism of missing data is analysed. Comparing the estimation results of IV(PP) and IV(PP) cc gives an impression of the magnitude of bias that is caused by MNAR versus no NAs and MCAR when applying a complete case analysis, as done by most of the construction methods. For both data generation strategies the results make clear that only IV ePP and IV ePP (rirs) are able to deal with non-ignorable missingness. All other methods exhibit bias. Even though, the estimation results of IV star are clearly biased, the F-statistic results indicate that this construction method will lead to a strong instrument. This is a meaningful illustration why testing the IV assumption and choosing a strong IV should not be the only consideration for this application and how important it is to investigate missingness in the data at hand.

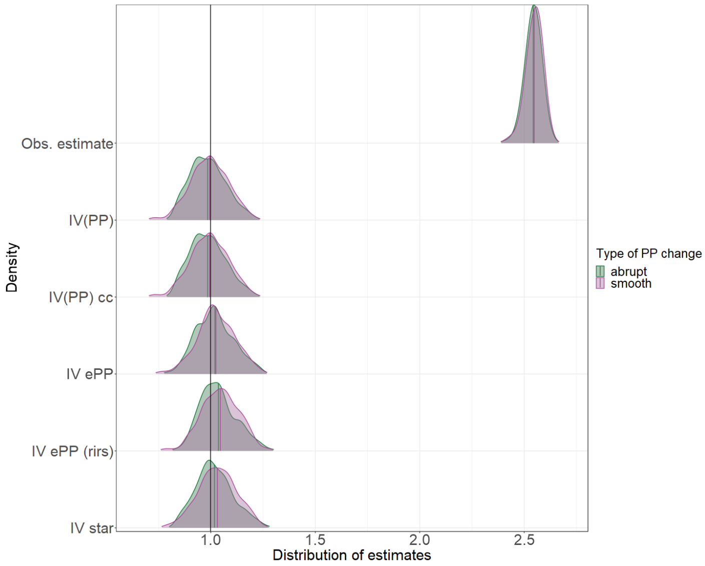


Figure 4.6: Estimation results of scenario 3: type of preference change. Estimation results for the data generation process using the Abrahamowicz model. Results are summarized for the model-based construction methods of Z*.*

The results of scenario 3 are given for the data generation strategy which simulates a PP variable as explained by Abrahamowicz et al. [93] only, as this simulation strategy offers the opportunity to simulate the change in preference as an abrupt or a smooth change. All previously explained scenarios are simulated with an abrupt change for this simulation strategy. The estimation results do not show much difference between abrupt and smooth change. Furthermore, the F-statistic results are consistent for both change types. This is not surprising for IV ePP and IV ePP (rirs) as the methods use a linear random effects model. But the results give confidence in the construction method by Abrahamowicz et al. [93]. IV star seems to be capable to find an adequate proxy for PP in case of an abrupt and a smooth change in preference.

	Scenario 1			Scenario 2			Scenario 3	
	Provider size (n_j)			Missing mechanism			Type of PP change	
	24	108	408	no NAs	MCAR	MNAR	abrupt	smooth
IV(PP)	27.5	116.94	451.4	451.4	275.46	238.62	451.4	315.26
IV(PP) cc	27.5	116.94	451.4	451.4	275.46	238.62	451.4	315.26
IV ePP	70.98	94.56	195.37	195.37	112.87	83.84	195.37	234.01
IV ePP (rirs)	60.5	155.12	410.79	410.79	243.22	186.92	410.79	398.32
IV star	25.2	68.51	127.99	127.99	103.04	104.31	127.99	151.95

Table 4.3: F-statistic results for the instrument Z *from the first stage regression model of the TSLS approach. The results are summarized for all scenarios. For this simulation the treatment decision* X *with the Abrahamowicz model. This table summarizes the results of all model-based construction methods for* Z*.*

	Scenario 1			Scenario 2		
	Provider size (n_j)			Missing mechanism		
	24	108	408	no NAs	MCAR	MNAR
IV(PP)	705.35	3181.47	12031.13	12031.13	7232.37	6790.5
IV(PP) cc	705.35	3181.47	12031.13	12031.13	7232.37	6790.5
IV ePP	65.53	22.41	20.64	20.64	6.24	3.13
IV ePP (rirs)	37.28	67.2	208.57	208.57	216.5	216.61
IV star	48.95	111.16	147.16	147.16	129.61	374.26

Table 4.4: F-statistic results for the instrument Z *from the first stage regression model of the TSLS approach. The results are summarized for all scenarios. For this simulation the treatment decision* X *is generated using the extended Ertefaie method. This table summarizes the results of all model-based construction methods for* Z*.*

With regards to the data availability, results of this simulation study show that the model-based approaches needed sufficient large amount of prescription data to estimate the treatment effect without bias. For smaller provider sizes, the rule-based methods were capable of estimating the treatment effect without bias. These results are discussed in more detail in Appendix 4.3. Only the construction method by the Ertefaie method and our extension method were able to adequately estimate the treatment effect in case of MNAR. Whereby, IV ePP showed only small bias but larger estimation variance compared to IV ePP (rirs). All model-based methods where able to produce treatment effects with small bias and acceptable estimation variance for most of the scenarios with a change in preference. The rule-based methods which use all prescription data within a provider to construct Z did struggle to estimate the treatment effect as they do not reflect on the change in PP, especially for small provider sizes. Additionally, the type of change did not seem to make much of a difference for the estimation performance for all construction methods.

As Uddin et al. [231] and our own simulation study have concluded, the validity of a preference-based IV strongly depends on the suitability of the data it is applied to. If possible, treatment effect estimates from multiple constructions should be derived and their coherence assessed. [231] In keeping with this spirit, in Section 4.5 we apply all of the rule and model-based construction methods discussed thus far to look at the comparative efficiency of two oral type 2 diabetes treatments.

4.5 Applied analysis: comparative effectiveness assessment of two treatments for type 2 diabetes

Type 2 diabetes (T2D) is a serious progressive metabolic disorder, characterized by hyperglycaemia and with an inherent risk of micro- and macrovascular complications. [114] Treatment mainly focuses on the control of blood glucose measured by the maintenance of glycated haemoglobin (HbA1c) levels. [126, 235] HbA1c level management is controlled by lifestyle changes and glucose-lowering agents. Two increasingly prescribed glucose-lowering agents are Sodium-glucose Co-transporter-2 Inhibitors (SGLT2i) and Dipeptidyl peptidase-4 Inhibitors. [126, 127] Although head-to-head RCT data suggest that the average glucose-lowering efficacy of both therapies is approximately similar [236], estimates are derived from highly selected cohorts which are not representative of the wider T2D population. This case study aimed to apply different IV methods in a real-world comparative effectiveness evaluation of the glucose-lowering efficacy of both therapies in an unselected T2D population. For this analysis different IVs for the preference of prescribing SGLT2i over DPP4i were constructed and applied. IV construction took place at the level of the GP practice in UK routine clinical data. UK routine clinical data contain key features of direct relevance to the methods discussed thus far, namely: (1) substantial between practice variation in preference for each drug even when patient covariates are taken into account; (2) evidence of a trend in prescription preference in favour of SGLT2i over time; and (3) missing data in key patient covariates that one would ideally like to adjust for. Figure 4.7 and 4.8 contain further information about the prescription trends between 2013-2020 and the prescription variation of SGLT2i between providers. Figure 4.7 highlights the increased prescribing of SGLT2i in recent years, which likely reflects their greater prominence in T2D treatment guidelines due to an accumulation of evidence on their cardiorenal benefits. Each circle in Figure 4.8 represents the proportion of SGLT2i prescriptions within a provider, relative to all other T2D oral agent prescriptions. Prescriptions of SGLT2i vary greatly between providers. As prescription of T2D agents is mostly done by primary care practices in the UK and not by specialised practices, a clustering of patients with specific characteristics is unlikely. This gives us confidence that a preference-based IV can be applied with this data. [80, 92] A similar study on the cardiovascular safety profile of Sulfonylureas using provider prescription preference with primary care data from Scotland was conduced by Wang et al. [237]. For this analysis Z was constructed using IV prevbpatient with $b = 10$ or all prescriptions of the previous 365 days.

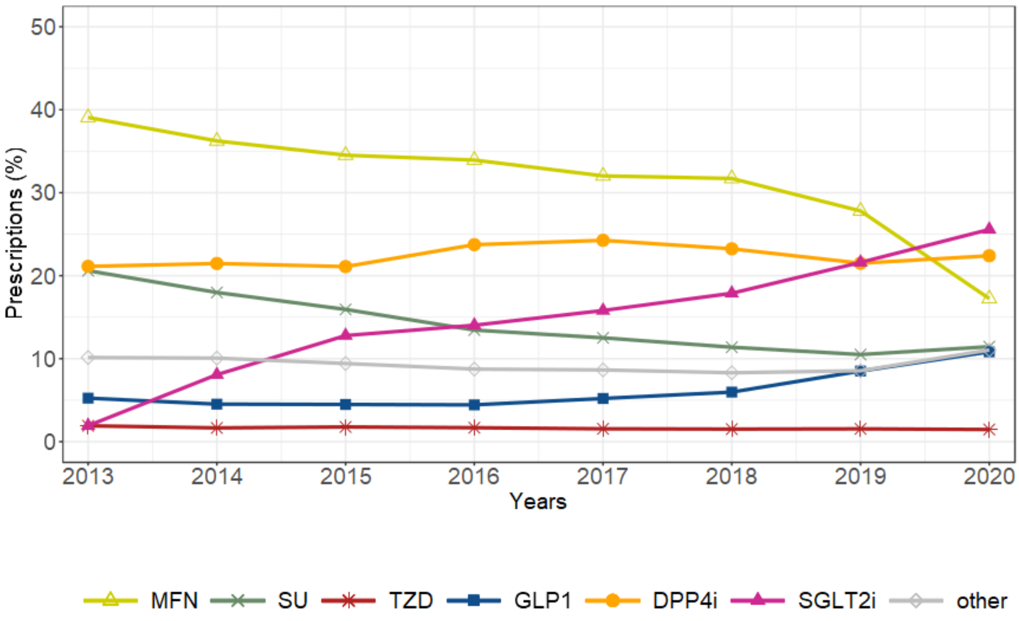


Figure 4.7: Prescribing trends of T2D oral agents in the study population data in the years 2013 to 2020. The trends are described by the yearly percentage of prescription of each agent respectively and relative to all T2D oral agent prescriptions.

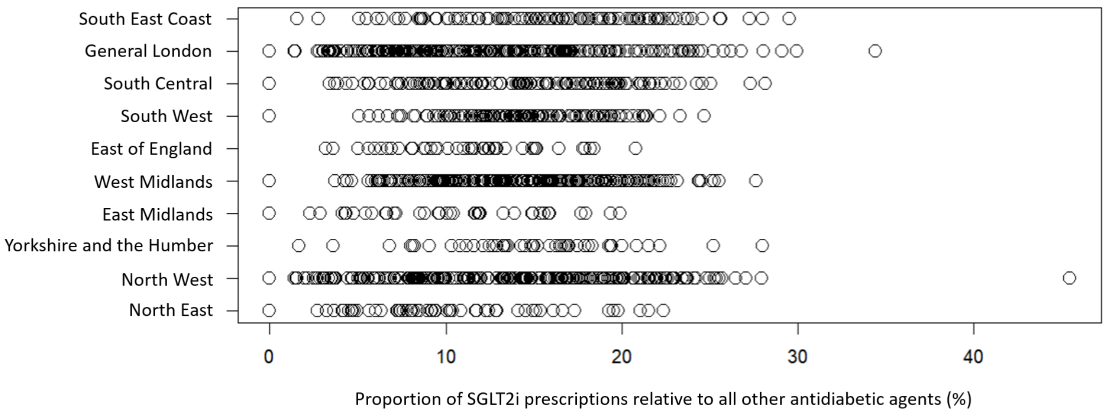


Figure 4.8: Prescription variation of SGLT2i between all practices in the study population. Each circle represents the proportion of SGLT2i relative to all prescriptions of T2D oral agents within a practice. The practices are clustered within their corresponding region for readability of the plot only.

4.5.1 Study Population and Data preparation

We used routine data from the Clinical Practice Research Datalink (CPRD) Aurum (download November 2021). [200] The study cohort of identified people with T2D included patients who initiated either SGLT2i ($N_{Tx} = 77229$) or DPP4i ($N_{Ct} = 109608$) between 2013 and 2020. A protocol on the identification strategy of people with T2D is explained in Rodgers et al. [203]. CPRD is a large source of primary healthcare data and encompasses 6.9% of the population in the UK. Furthermore, it is considered to be representative of the UK population regarding age, sex and ethnicity. [200] The study cohort comprised people with T2D initiating either SGLT2i ($N_{Tx} = 77229$) or DPP4i ($N_{Ct} = 109608$) between 2013 and 2020 who had baseline HbA1c 53-120 mmol/mol and had an estimated glomerular filtration rate (eGFR) $\geq 45 mL/min/1.73m^2$. The chosen HbA1c range represents the lower threshold for glucose-lowering medication initiation in clinical guidelines and for severe hyperglycemia. SGLT2i's were contraindicated in the UK in individuals with eGFR $< 45 mL/min/1.73m^2$ and not licensed for use below this threshold for the majority of the study period. CPRD data extraction followed our previously published protocol. [203] Baseline clinical characteristics (measured confounders) are reported in Table 4.5 for each drug arm, and were included in the outcome model of 12 month achieved HbA1c (mmol/mol), closest value to 12 months in the 9-15 months after treatment initiation, on unchanged therapy. Each practice treated on average 132 individuals with either drug over the study period. A small number of practices treated only 1 or 2 patients with either treatment and the maximum number of individuals treated by a practice was 1911.

Variable	DPP4i	SGLT2i
	$N_{Ct} = 109608$	$N_{Tx} = 77229$
HbA1c (mmol/mol)	73.2 (14.1)	77.2 (14.8)
BMI (kg/m^2)	31.9 (6.58)	33.8 (6.82)
eGFR (ml/min/1.73m^2)	87.7 (18.8)	94.8 (15.1)
ALT (U/L)	31.7 (19.4)	34.5 (20.1)
Age (years)	63.0 (12.5)	58.3 (10.5)
T2D duration (years)	9.01 (6.54)	9.84 (6.34)

Variable	DPP4i $N_{Ct} = 109608$	SGLT2i $N_{Tx} = 77229$
Sex		
female	43973 (40.1%)	30022 (38.9%)
male	65635 (59.9%)	47207 (61.1%)
Year of treatment prescription		
2013	12651 (11.54%)	1254 (1.62%)
2014	13038 (11.9%)	5537 (7.17%)
2015	14832 (13.53%)	10155 (13.15%)
2016	16704 (15.24%)	11121 (14.4%)
2017	16960 (15.47%)	12398 (16.05%)
2018	16326 (14.89%)	14325 (18.55%)
2019	13969 (12.74%)	15862 (20.54%)
2020	5128 (4.68%)	6577 (8.52%)
Ethnicity		
White	84068 (76.7%)	59393 (76.9%)
South Asian	15026 (13.7%)	10963 (14.2%)
Black	5922 (5.4%)	3352 (4.3%)
Other	1647 (1.5%)	1148 (1.5%)
Mixed	1063 (1.0%)	775 (1.0%)
Deprivation	5.91 (2.85)	5.89 (2.86)
Smoking status		
Active smoker	17771 (16.2%)	12744 (16.5%)
Ex-smoker	58610 (53.5%)	41860 (54.2%)
Non-smoker	27859 (25.4%)	19451 (25.2%)
Number of concurrent treatments		
1	11253 (10.3%)	4293 (5.6%)
2	63519 (58.0%)	33783 (43.7%)
3+	34836 (31.8%)	39153 (50.7%)

Variable		DPP4i $N_{Ct} = 109608$	SGLT2i $N_{Tx} = 77229$
Line of treatment			
	1	2461 (2.2%)	571 (0.7%)
	2	43887 (40.0%)	14628 (18.9%)
	3	44498 (40.6%)	21041 (27.2%)
	4+	18762 (17.1%)	40989 (53.1%)
Patients ever taken Insulin			
	yes	5312 (4.85%)	10691 (13.84%)

Table 4.5: Baseline characteristics of the CPRD T2D cohort for patients starting DPP4i ($N_{Ct} = 109608$*) or SGLT2i (*$N_{Tx} = 77229$*) after 2013. Values are shown in mean (standard deviation) unless otherwise stated. Abbreviations: HbA1c (glycated haemoglobin), BMI (body mass index), eGFR (estimated glomerular rate), measured using the CKD-EPI Creatinine equation (2021), ALT (alanine aminotransferase), T2D (type 2 diabetes). Furthermore, deprivation was measured using the English Index of Multiple Deprivation (IMD) decile (1=most deprived, 10=least deprived).*

Further data preparation of the study population was needed in order to apply all construction methods of Z and the TSLS IV estimation approach. For all construction methods a complete case dataset without missingness on the outcome variable was required. Additionally, for all construction methods other than the Ertefaie method and its extension method, a complete case dataset on the measured confounders was essential. In Table 4.18 in Appendix 4.6 an overview of the structure of missing values in the study population is given. Each construction method requires a different minimum number of patients treated by each provider ($n_{j,min}$). Providers with too little data are excluded from the analysis. A summary on $n_{j,min}$ together with information about the dataset sizes after exclusion of too small providers is given in Appendix 4.6 in Table 4.19.

4.5.2 Results

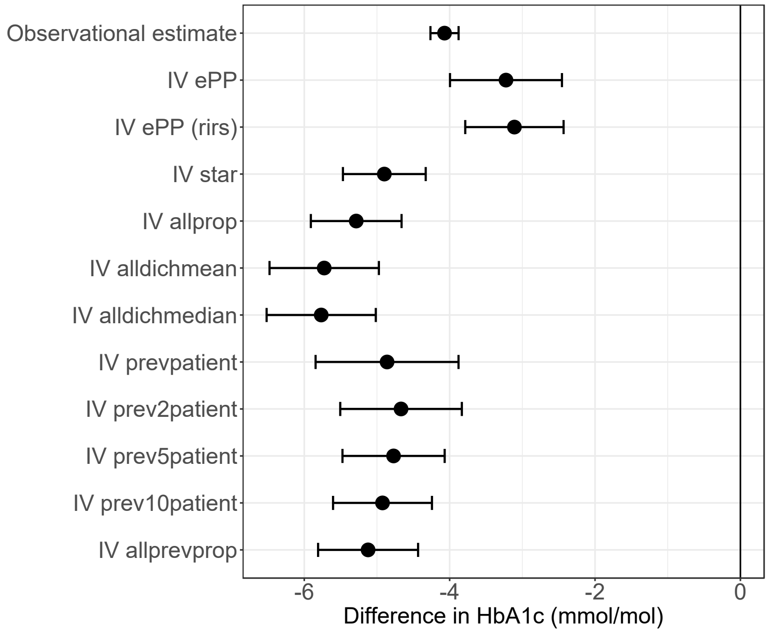


Figure 4.9: Estimation results of the relative treatment effect of SGLT2i versus DPP4i on the reduction of HbA1c (mmol/mol). Values smaller than 0 indicate that SGLT2i has a stronger HbA1c decreasing effect compared to DPP4i. Results are shown for a multivariable regression analysis (observational estimate) and all IV estimates employing the construction methods of a preference-based IV.

It was possible to apply all construction methods for the instrument explained in Section 4.3 in this application case study. Slightly different subsets of the study population were needed for each construction method due to different requirements on complete case data and minimum number of patients treated by each provider ($\mathsf{n_{j,min}}$). From Table 4.6 it is clear that IV prev10patient used the smallest dataset for the analysis as this method requires a complete case dataset on the outcome variable and measured confounders and requires $\mathsf{n_{j,min}} > 11$. Additionally, the data of the first 10 patients treated by each practice were excluded from the IV estimation as Z cannot be calculated for these patients.

The estimation results for the application case study are given in Figure 4.9 together with 95% confidence intervals (CIs). CIs are taken from the outcome model, this ignores the uncertainty in the first stage model. All approaches show a consistently greater HbA1c reduction with SGLT2i compared to DPP4i. Compared with the other approaches, this difference is attenuated using the Ertefaie method and our extension method. From the simulation outlined in Section 4.4, we have seen that we can trust these methods to account for non-ignorable missingness. As in the simulation study we see that our proposed method estimates the treatment effect with slightly more efficiency compared to IV ePP. All other methods shown in this plot rely on complete case analysis and therefore use smaller and potentially more selective datasets regarding patient characteristics. Using complete case analysis can lead to bias if the missingness is MNAR. [45, 97, 230] Hence, it is possible that by using complete case analysis we exclude patients with certain characteristics and therefore overestimate the relative treatment effect. In Appendix 4.7 the estimation results for the case study are given for which all IV construction methods are applied to the same complete case dataset with $\mathsf{n}_{\mathsf{j,min}} > 11$. The results show that IV ePP and IV ePP (rirs) still lead to significantly smaller relative blood glucose benefit estimate compared to all of IV construction methods. Additionally, the Ertefaie method and its extension method utilize slightly different outcome models to estimate the treatment effect (second stage model of the IV estimation) because both methods only include confounders which are measured for all patients ($\mathbf{W_{obs}}$, as explained in Section 4.3.3). As the estimation results for IV ePP and IV ePP (rirs) are in agreement, this might indicate that changing preferences of providers is less of an issue of concern in this analysis or that provider sizes are large enough for IV ePP to on average reflect suitably on PP.

4.6 Discussion

In this paper we conducted a state of the art performance analysis of the known construction methods for a preference-based instrument. With this study we add to the already existing literature on the performance evaluation of preference-based IVs [82, 93, 97, 231] by giving a comprehensive overview over all construction methods and evaluating all methods with respect to three important aspects: availability of prescription data within a provider, different missing data mechanisms for missing data in measured confounders and change in provider preference over time. Additionally, we proposed an extended version of the construction method by Ertefaie et al. [97] which aimed to combined the ability to deal with non-ignorable missingness and change in PP using a random intercept and random slope model for the construction of Z. A simulation study was conducted using two different data generation strategies, to evaluate the performance independent of a specific generation process for X that might benefit certain methods. Furthermore, all construction methods were showcased with a real life primary care dataset in a relative effectiveness study of two T2D oral agents. This case study outlines which data requirements are needed for each method to be applied to real life data, such as the necessity for complete case analysis, the exclusion of data due to insufficient prescription data or the inability to calculate a value for the proxy instrument Z with a specific construction method.

Our results indicate that most model-based and rule-based construction methods do not have substantial problems in accounting for a change in provider preference. An exception from this are the rule-based methods which utilized all prescription information within a provider (IV allprop and its variations) and cannot reflect on change in preference appropriately. Our extension method models preference change with a random intercept random slope model. Results of the simulation study indicate that the method is able to estimate the treatment effect without bias. Additionally, in the simulation study and application case study the extension method estimated the treatment effect more efficiently compared to its original version proposed by Ertefaie et al. [97]. Especially the performance of the more complex model-based approaches depend on the availability of sufficient data per provider. All model-based methods struggled to estimate the treatment effect without bias in case of small provider sizes. When applying a model-based method to real life data the available data within providers is therefore a crucial consideration. More simple rule-based methods such as IV prevpatient proposed by Brookhart et al. [92] could be considered as alternative in case of small datasets. The Ertefaie method and our extension method are capable of estimating the true treatment effect even in case of non-ignorable missingness in measured confounders. This makes them favorable in many observational research studies. Both construction methods will still require the use of complete case datasets based on the outcome variable, which may also lead to a selection of patients with specific characteristics and therefore a distortion of the treatment effect. All in all, the application case study showed the usefulness of triangulating results from different construction methods as proposed by Uddin et al. [231]. In doing so non-consistent results between the IV estimates can be discussed and sources of bias in the respective study may be discovered. [9]

This study has some limitations which opens possibilities for further research. We have only tested the Ertefaie method and our extension on one specific mechanism to generate MNAR. With this, missingness was created based on the outcome variable value, the missing value itself, an unmeasured confounder and a provider level influence on missingness. We used the same generation process for missingness as in the original paper [97]. It would be valuable to test both methods on different selection models for MNAR missingness to verify our findings and to explore sensitivity to miss-specification of the missingness model. Results of the simulation for the rule-based methods are presented in Appendix 4.3 and 4.4. When applying the construction methods to datasets with different provider sizes, the F-statistic results show that IV prevpatient and its variations are weak instruments with F-statistic values smaller than 10. Acceptable IV strength was only achieved by considering 5 or more previous prescription in the IV construction and for large provider sizes of $n_j = 408$. Interestingly, the estimation results for these construction methods did not show weak instrument bias (Figure 4.12) as we would have expected. Further investigations will help to understand if the F-statistic results might be misleading for this instrument, maybe due an introduction of serial autocorrelation between Z and X when using previous prescriptions to reflect on provider preference. For the application case study we faced some data limitation using CPRD primary care data on the information available to construct a proxy for provider preference. Only information on the allocation of patients to practices was used in this analysis as the Staff ID variable provided in CPRD might not always reflect on the prescribing practitioner but the person entering the data. In reality, patients will be treated by different physicians within a practice with different prescription preference. By constructing a proxy on practice level, information of prescription pattern will be aggregated which will lead to measurement bias unless all physicians within a practice have the same preference.

In summary, our study shows that IV methods using provider preference can be a useful tool for causal inference from observational health data, with both model-based and rule-based construction methods of preference-based instruments performing well in our simulation study as long as changes in provider preference over time are incorporated. Both the Ertefaie method and our proposed extension method are capable of estimating causal treatment effects even in case of non-ignorable missingness in measured confounders, and are recommended where sufficient data are available.

Acknowledgements

This article is based in part on data from the Clinical Practice Research Datalink obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. CPRD data is provided by patients and collected by the NHS as part of their care and support. This study was supported by the National Institute for Health and Care Research Exeter Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Funding

Approval for the study was granted by the CPRD Independent Scientific Advisory Committee (ISCA 22_002000). LMG, JB and JMD are supported by the UK Research and Innovation Expanding Excellence in England fund. BMS and JMD receive funding from MRC and MASTERMIND. This research was funded by the Medical Research Council (UK) (MR/N00633X/1).

Conflict of interest

JB is a part time employee of Novo Nordisk. This project is unrelated to his work for the company.

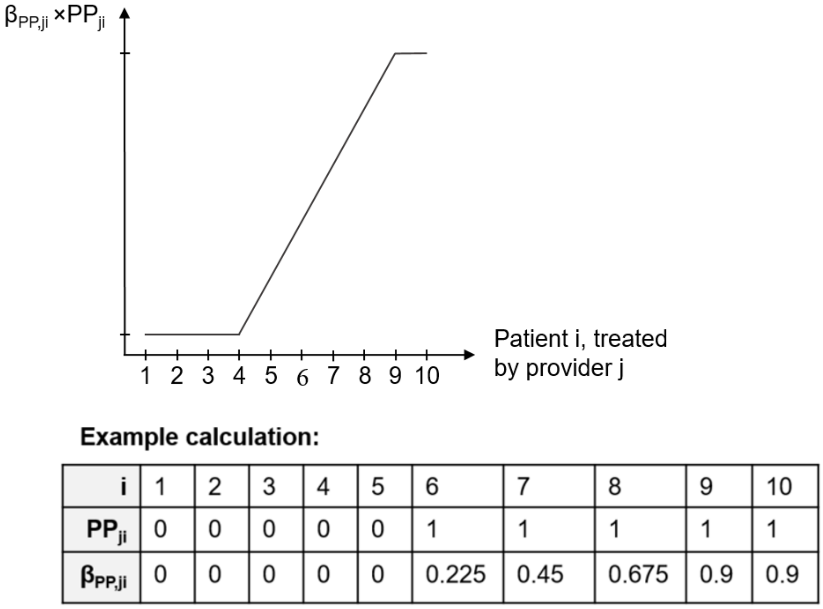
Data availability statement

Data from CPRD is available to all researchers following successful application to the ISAC. Source code for this research for all simulations and the application study in this paper is available at `https://github.com/GuedemannLaura/ppIV`.

4.7 Appendices

Appendix 4.1 Calculation example for the data generation of a smooth change in prescribing preference

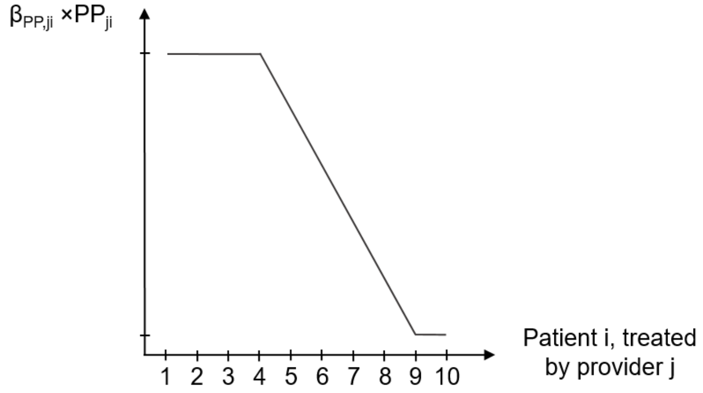
Figure 4.10 and 4.11 show calculation examples of how the smooth change was simulated based on Abrahamowicz et al. [93] and applied for the simulation study explained in Section 4.4. Smooth changes for a provider j changing preference form treatment A to B and vice versa are shown. In these examples provider j treats $n_j = 10$ patients, and changes preference after the patient $i^\star = 5$ was treated. The change in preference takes place for $L_j = 4$ periods/ treated patients and $PP_{final,ji} = PP_{initial,ji} = 0.9$. The influence of the change on the treatment decision $\beta_{PP,ji}$ during the change period is calculated with formulas (4.7) and (4.8) respectively, as explained in Section 4.4.



Example calculation:

i	1	2	3	4	5	6	7	8	9	10
PP_{ji}	0	0	0	0	0	1	1	1	1	1
$\beta_{PP,ji}$	0	0	0	0	0	0.225	0.45	0.675	0.9	0.9

Figure 4.10: Simulation of a smooth change for a provider j changing preference from treatment A to B.



Example calculation:

i	1	2	3	4	5	6	7	8	9	10
PP_{ji}	1	1	1	1	1	1	1	1	0	0
$\beta_{PP,ji}$	0.9	0.9	0.9	0.9	0.9	0.675	0.45	0.225	0	0

Figure 4.11: Simulation of a smooth change for a provider j *changing preference from treatment B to A.*

Appendix 4.2 Additional information on the data generation process

Table 4.6 shows additional information on the data generation process for the simulation study explained in Section 4.4. The left side of this table shows the variance of the simulated outcome variable Y and the proportion of patients treated with $X_{ji} = 1$ (in %) for the simulation strategy using the Abrahamowicz model to generate X. The right side of the table shows these results for the simulation which employs the extended Ertefaie model to simulate X.

		Simulation strategy for X			
		Abrahamowicz model		Extended Ertefaie model	
		Var(Y)	$p(X_{ji} = 1) \times 100$	Var(Y)	$p(X_{ji} = 1) \times 100$
Scenario 1	$n_j = 24$	7.798	42.788	6.996	56.394
	$n_j = 108$	7.778	42.767	6.992	55.825
	$n_j = 408$	7.782	42.462	6.991	56.085
Scenario 2	no NAs	7.782	42.462	6.991	56.085
	MCAR	7.783	43.74	6.994	56.302
	MNAR	7.783	43.732	6.994	56.301
Scenario 3	abrupt	7.782	42.462		
	smooth	7.776	41.748		

Table 4.6: Additional information on the outcome variance and the proportion of treated patients for both data generation strategies employed in the simulation explained in Section 4.4.

Appendix 4.3 Simulation results for rule-based construction methods

In Section 4.4 a state of the art simulation study is described which focuses on three important aspects to consider when using provider preference-based IVs: the availability of prescription data for each provider of the study population, missing data in the measured confounders and possible change in provider preference over time. Results summary outlined in Section 4.4 focused on the model-based construction methods of Z. The additional results for all rule-based methods introduced in Section 4.3 are given below. Table 4.7 summarized the rule-based IV construction methods and their abbreviations.

Abbreviation	Method
IV allprop	IV based on all prescriptions (proportion)
IV alldichmean	IV based on all prescriptions (dichotomized with mean)
IV alldichmedian	IV based on all prescriptions (dichotomized with median)
IV prevpatient	IV based on previous prescription
IV prev2patient	IV based on previous 2 prescriptions
IV prev5patient	IV based on previous 5 prescriptions
IV prev10patient	IV based on previous 10 prescriptions
IV allprevprop	IV based on all previous prescriptions

Table 4.7: Summary of rule-based construction methods applied in the simulation and their abbreviations.

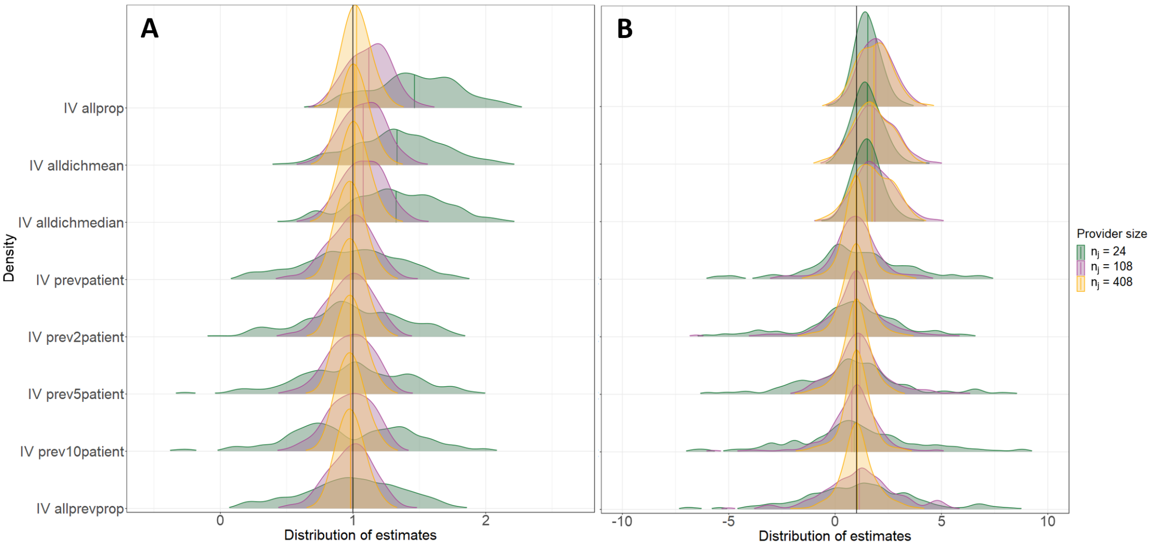


Figure 4.12: Estimation results of scenario 1: change in provider size. Panel A: estimation results for the data generation process using the Abrahamowicz model. Panel B: estimation results for the data generation process using the extended Ertefaie model to simulate change in preference. Results are summarized for the rule-based construction methods of Z.

For scenario 1 and in Panel A, the construction methods which use a subset of previous prescriptions are unbiased for larger n_j as they can reflect on a change in PP. We do not see and improvement of the estimation variance when including more previous prescription data in the construction of Z. But the F-statistic results in Table 4.8 reflects that including more previous patient in the construction of Z leads to stronger instruments. Similar results have been shown by Uddin et al. [231]. They concluded that IV prev10patient outperformed IV prev5patient and IV prevpatient with regards to strength. In contrast, IV allprop and its variations cannot reflect on PP because all prescription data within a provider is used simultaneously for the construction of Z. The estimation results are biased for $n_j = 24$ and $n_j = 108$. In Panel B, the results for the simulation strategy using the extended Ertefaie model to simulate X are given. The results show generally larger estimation variance for methods using a subset of previous prescriptions and for smaller n_j. For larger n_j the estimation variance look similar over all methods. As for the first simulation strategy, methods using all prescription information to construct Z are biased even for larger n_j. This results is also reflected in their coverage rates summarized in Table 4.13.

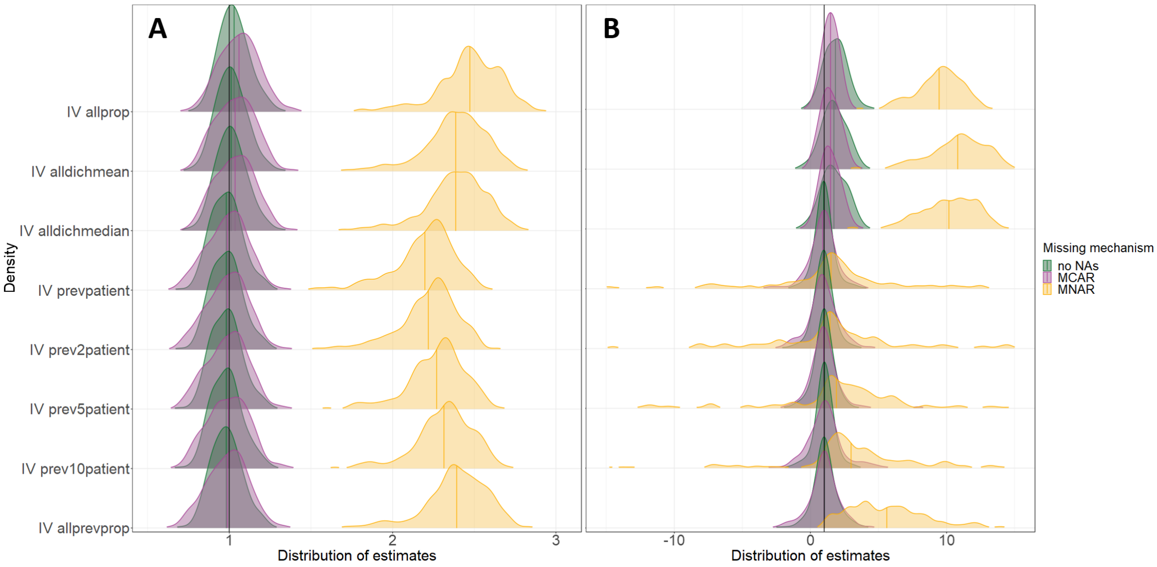


Figure 4.13: Estimation results of scenario 2: missing data mechanisms. Panel A: estimation results for the data generation process using the Abrahamowicz model. Panel B: estimation results for the data generation process using the extended Ertefaie model to simulate change in preference. Results are summarized for the rule-based construction methods of Z.

For scenario 2, all model-based construction methods result in biased treatment effect estimations in case of non-ignorable missingness. These results hold for both stimulation strategies.

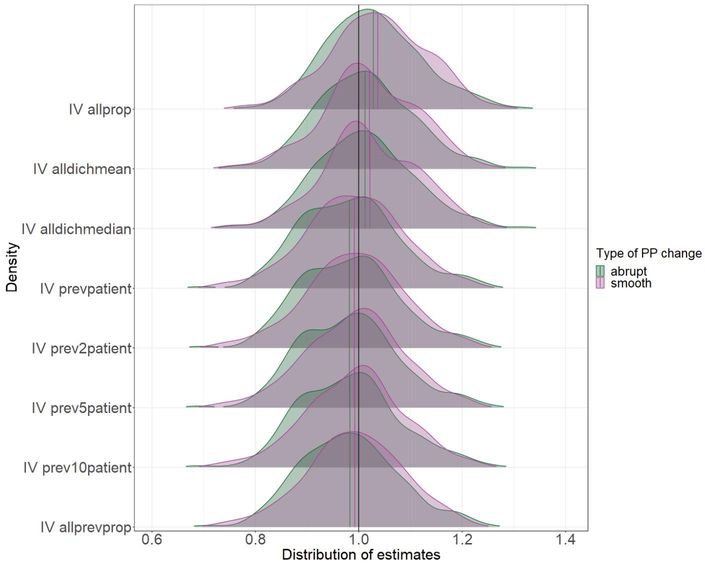


Figure 4.14: Estimation results of scenario 3: type of preference change. Estimation results for the data generation process using the Abrahamowicz model. Results are summarized for the rule-based construction methods of Z.

For scenario 3, the estimation results for the rule-based methods are consistent for both types of change, similarly to the results of the model-based construction methods. As in scenario 1, we see bias for IV allprop and its variations as they do not account for a change in preference at all, but the bias is similar for both change types.

	Scenario 1			Scenario 2			Scenario 3	
	Provider size (n_j)			Missing mechanism			Type of PP change	
	24	108	408	no NAs	MCAR	MNAR	abrupt	smooth
IV allprop	63.35	95.81	234.39	234.39	160.32	93.49	234.39	274.52
IV alldichmean	42.42	62.81	158.18	158.18	107.13	48.24	158.18	198.31
IV alldichmedian	41.73	62.02	155.99	155.99	105.81	50.05	155.99	192.56
IV prevpatient	1.01	1.43	3.78	3.78	2.8	9.9	3.78	3.43
IV prev2patient	1.38	2.15	6.47	6.47	4.15	17.15	6.47	5.56
IV prev5patient	1.39	3.96	14.26	14.26	8.96	32.57	14.26	11.8
IV prev10patient	1.25	6.24	26.96	26.96	17.03	46.32	26.96	22.16
IV allprevprop	1.21	5.89	54.35	54.35	24.26	63.68	54.35	66.06

Table 4.8: F-statistic results for the instrument Z *from the first stage regression model of the TSLS approach. The results are summarized for all scenarios. For this simulation the treatment decision* X *using the Abrahamowicz model. This table summarized the results of all rule-based construction methods for* Z*.*

	Scenario 1			Scenario 2		
	Provider size (n_j)			Missing mechanism		
	24	108	408	no NAs	MCAR	MNAR
IV allprop	75.26	29.05	26.43	26.43	52.03	93.93
IV alldichmean	51.7	18.79	17	17	33.91	62.63
IV alldichmedian	50.16	18.4	16.8	16.8	33.55	63.38
IV prevpatient	8.2	44.95	186.92	186.92	46.36	62.37
IV prev2patient	9.05	45.77	190.14	190.14	41.15	60.2
IV prev5patient	8.12	44.32	156.08	156.08	39.72	65.6
IV prev10patient	7.42	37.61	136.51	136.51	39.07	74.18
IV allprevprop	3.21	10.25	90.62	90.62	39.61	213.96

Table 4.9: F-statistic results for the instrument Z *from the first stage regression model of the TSLS approach. The results are summarized for all scenarios. For this simulation the treatment decision* X *is generated using the extended Ertefaie model. This table summarized the results of all rule-based construction methods for* Z*.*

Appendix 4.4 Additional estimation performance results for all construction methods and all simulation scenarios

In the following additional estimation performance measures for the simulation study are summarized. Results on the bias, standard error (SE) coverage (in %) and root mean squared error (RMSE) are given for all scenarios and proxy measure construction methods.

Method	n_j	Bias	SE	Coverage	RMSE
Obs. estimate	24	1.5552	0.0083	0	1.5596
	108	1.545	0.0041	0	1.5461
	408	1.5427	0.0026	0	1.5431
IV(PP)	24	-0.0119	0.0236	98.5	0.3331
	108	-0.0173	0.0109	98	0.1548
	408	-0.0149	0.0061	97.5	0.0867
IV(PP) cc	24	-0.0119	0.0236	98.5	0.3331
	108	-0.0173	0.0109	98	0.1548
	408	-0.0149	0.0061	97.5	0.0867
IV ePP	24	0.4321	0.0258	74	0.5649
	108	0.1196	0.011	96.5	0.1957
	408	0.0207	0.0065	95.5	0.0942
IV ePP (rirs)	24	0.3987	0.0285	75.5	0.5664
	108	0.1681	0.0111	84	0.2298
	408	0.0355	0.0061	94.5	0.0933
IV star	24	0.2462	0.0248	93	0.4281
	108	0.1111	0.0119	93.5	0.2007
	408	0.0169	0.0064	96	0.0918
IV allprop	24	0.4641	0.0228	71	0.5649
	108	0.12	0.0111	95.5	0.1968
	408	0.0287	0.0064	94.5	0.0948
IV alldichmean	24	0.3329	0.0241	87	0.476
	108	0.0766	0.0112	96.5	0.1759
	408	0.0119	0.0066	95	0.0932

Method	n_j	Bias	SE	Coverage	RMSE
IV alldichmedian	24	0.3259	0.0243	87.5	0.473
	108	0.0758	0.0113	96.5	0.1759
	408	0.0119	0.0066	95	0.0938
IV prevpatient		0.0022	0.0265	98.5	0.3735
	108	-0.0166	0.0119	97.5	0.1691
	408	-0.0179	0.0067	97.5	0.0964
IV prev2patient	24	-0.0014	0.0272	98	0.3835
	108	-0.0178	0.012	97.5	0.1697
	408	-0.0182	0.0067	98	0.0959
IV prev5patient	24	0.0001	0.0302	97	0.4256
	108	-0.0166	0.0120	98	0.1699
	408	-0.0185	0.0067	98	0.0957
IV prev10patient	24	-0.002	0.0334	97	0.4716
	108	-0.0171	0.0123	98	0.1737
	408	-0.0175	0.0067	97	0.0965
IV allprevprop	24	0.0012	0.0263	98	0.3717
	108	-0.0154	0.0117	97.5	0.1664
	408	-0.0173	0.0066	97.5	0.0949

Table 4.10: Summary of performance measures for scenario 1. For this simulation the treatment decision X *is generated using the Abrahamowicz model.*

Method	Missing mechanism	Bias	SE	Coverage	RMSE
obs. estimate	no NAs	1.5427	0.0026	0	1.5431
	MCAR	1.5508	0.0029	0	1.5513
	MNAR	1.6638	0.0032	0	1.6644
IV(PP)	no NAs	-0.0149	0.0061	97.5	0.0867
	MCAR	-0.0143	0.0063	96	0.0893
	MNAR	-0.0143	0.0063	0	0.0893
IV(PP) cc	no NAs	-0.0149	0.0061	97.5	0.0867
	MCAR	-0.0109	0.0082	96	0.1157
	MNAR	1.0428	0.013	96	1.0588

Method	Missing mechanism	Bias	SE	Coverage	RMSE
IV ePP	no NAs	0.0207	0.0065	95.5	0.0942
	MCAR	0.0712	0.0182	94.5	0.2667
	MNAR	0.0402	0.0191	95	0.2729
IV ePP (rirs)	no NAs	0.0355	0.0061	94.5	0.0933
	MCAR	0.1206	0.0157	95.5	0.2525
	MNAR	0.1063	0.0167	95	0.2586
IV star	no NAs	0.0169	0.0064	96	0.0918
	MCAR	0.0518	0.009	95	0.1373
	MNAR	1.4479	0.0137	0	1.4607
IV allprop	no NAs	0.0287	0.0064	94.5	0.0948
	MCAR	0.0601	0.0085	93	0.1342
	MNAR	1.4727	0.0137	0	1.4853
IV alldichmean	no NAs	0.0119	0.0066	95	0.0932
	MCAR	0.0349	0.0085	97	0.1253
	MNAR	1.386	0.0134	0	1.3988
IV alldichmedian	no NAs	0.0119	0.0066	95	0.0938
	MCAR	0.0339	0.0085	97	0.1248
	MNAR	1.3856	0.0134	0	1.3985
IV prevpatient	no NAs	-0.0179	0.0067	97.5	0.0964
	MCAR	-0.0117	0.0089	95	0.1256
	MNAR	1.1966	0.0132	0	1.2109
IV prev2patient	no NAs	-0.0182	0.0067	98	0.0959
	MCAR	-0.0112	0.0089	95.5	0.1257
	MNAR	1.2198	0.013	0	1.2335
IV prev5patient	no NAs	-0.0185	0.0067	98	0.0957
	MCAR	-0.0112	0.0089	95	0.1261
	MNAR	1.2689	0.0126	0	1.2813
IV prev10patient	no NAs	-0.0175	0.0067	97	0.0965
	MCAR	-0.0078	0.009	96.5	0.1265
	MNAR	1.3157	0.0128	0	1.328

Method	Missing mechanism	Bias	SE	Coverage	RMSE
IV allprevprop	no NAs	-0.0173	0.0066	97.5	0.0949
	MCAR	-0.0112	0.0089	95.5	0.1258
	MNAR	1.393	0.0134	0	1.4058

Table 4.11: Summary of performance measures for scenario 2. For this simulation the treatment decision X *is generated using the Abrahamowicz model.*

Method	Change type	Bias	SE	Coverage	RMSE
obs. estimate	abrupt	1.5427	0.0026	0	1.5431
	smooth	1.5483	0.0026	0	1.5487
IV(PP)	abrupt	-0.0149	0.0061	97.5	0.0867
	smooth	-0.0068	0.0064	97.5	0.0903
IV(PP) cc	abrupt	-0.0149	0.0061	97.5	0.0867
	smooth	-0.0068	0.0064	97.5	0.0903
IV ePP	abrupt	0.0207	0.0065	95.5	0.0942
	smooth	0.0255	0.0064	95	0.0941
IV ePP (rirs)	abrupt	0.0355	0.0061	94.5	0.0933
	smooth	0.046	0.0063	92	0.1002
IV star	abrupt	0.0169	0.0064	96	0.0918
	smooth	0.0318	0.0065	95	0.0968
IV allprop	abrupt	0.0287	0.0064	94.5	0.0948
	smooth	0.0371	0.0065	93	0.0988
IV alldichmean	abrupt	0.0119	0.0066	95	0.0932
	smooth	0.0207	0.0064	97	0.0931
IV alldichmedian	abrupt	0.0119	0.0066	95	0.0938
	smooth	0.0213	0.0065	96.5	0.0936
IV prevpatient	abrupt	-0.0179	0.0067	97.5	0.0964
	smooth	-0.009	0.0066	95	0.0936
IV prev2patient	abrupt	-0.0182	0.0067	98	0.0959
	smooth	-0.0086	0.0066	95	0.0929
IV prev5patient	abrupt	-0.0185	0.0067	98	0.0957
	smooth	-0.0083	0.0066	94.5	0.0936

Method	Change type	Bias	SE	Coverage	RMSE
IV prev10patient	abrupt	-0.0175	0.0067	97	0.0965
	smooth	-0.0076	0.0066	95.5	0.0936
IV allprevprop	abrupt	-0.0173	0.0066	97.5	0.0949
	smooth	-0.0075	0.0065	9	0.0924

Table 4.12: Summary of performance measures for scenario 3. For this simulation the treatment decision X *is generated using the Abrahamowicz model.*

Method	n_j	Bias	SE	Coverage	RMSE
obs. estimate	24	0.3568	0.0063	3.5	0.3677
	108	0.3664	0.0039	0	0.3705
	408	0.3616	0.0027	0	0.3637
IV(PP)	24	-0.0191	0.008	96	0.1138
	108	-0.0039	0.0039	94.5	0.0549
	408	-0.009	0.002	93.5	0.0303
IV(PP) cc	24	-0.0191	0.008	96	0.1138
	108	-0.0039	0.0039	94.5	0.0549
	408	-0.009	0.002	93.5	0.0303
IV ePP	24	0.2263	0.3708	92.5	5.235
	108	0.293	0.2243	93	3.1779
	408	0.1837	0.1109	90.5	1.5754
IV ePP (rirs)	24	0.1751	0.1099	97	1.5605
	108	0.1957	0.0726	94	1.0433
	408	0.0105	0.044	96.5	0.6201
IV star	24	0.3880	0.0538	95	0.8521
	108	0.3638	0.0553	88.5	0.8609
	408	0.3506	0.0442	86	0.7155
IV allprop	24	0.5342	0.0394	82	0.7713
	108	0.9229	0.0547	78	1.2028
	408	0.8371	0.0582	78.5	1.1725
IV alldichmean	24	0.524	0.049	89.5	0.8677
	108	0.8589	0.0681	86.5	1.2882
	408	0.7379	0.0634	85.5	1.1597

Method	n_j	Bias	SE	Coverage	RMSE
IV alldichmedian	24	0.5228	0.0483	89.5	0.8583
	108	0.874	0.0676	87	1.2934
	408	0.7461	0.0652	84	1.1842
IV prevpatient	24	0.0733	0.2228	95	3.1432
	108	0.0188	0.0873	96	1.2324
	408	-0.0135	0.0529	96	0.7461
IV prev2patient	24	0.071	0.2268	95.5	3.2
	108	-0.088	0.1119	95	1.5803
	408	-0.017	0.0595	94.5	0.8392
IV prev5patient	24	0.1868	0.3076	96.5	4.3437
	108	0.1625	0.0952	96.5	1.3533
	408	-0.0259	0.0513	92.5	0.7237
IV prev10patient	24	0.2317	0.4096	96	5.7833
	108	-0.1494	0.1164	96.5	1.6494
	408	0.0448	0.0558	93.5	0.7888
IV allprevprop	24	-0.0751	0.2498	98	3.525
	108	0.2378	0.158	96.5	2.2412
	408	0.0382	0.0677	94	0.9563

Table 4.13: Summary of performance measures for scenario 1. For this simulation the treatment decision X *is generated using the extended Ertefaie model.*

Method	Missing mechanism	Bias	SE	Coverage	RMSE
obs. estimate	no NAs	0.3568	0.0063	3.5	0.3677
	MCAR	0.3664	0.0039	0	0.3705
	MNAR	0.3616	0.0027	0	0.3637
IV(PP)	no NAs	-0.0191	0.008	96	0.1138
	MCAR	-0.0039	0.0039	94.5	0.0549
	MNAR	-0.009	0.002	93.5	0.0303
IV(PP) cc	no NAs	-0.0191	0.008	96	0.1138
	MCAR	-0.0039	0.0039	94.5	0.0549
	MNAR	-0.009	0.002	93.5	0.0303

Method	Missing mechanism	Bias	SE	Coverage	RMSE
IV ePP	no NAs	0.2263	0.3708	92.5	5.235
	MCAR	0.293	0.2243	93	3.1779
	MNAR	0.1837	0.1109	90.5	1.5754
IV ePP (rirs)	no NAs	0.1751	0.1099	97	1.5605
	MCAR	0.1957	0.0726	94	1.0433
	MNAR	0.0105	0.044	96.5	0.6201
IV star	no NAs	0.388	0.0538	95	0.8521
	MCAR	0.3638	0.0553	88.5	0.8609
	MNAR	0.3506	0.0442	86	0.7155
IV allprop	no NAs	0.5342	0.0394	82	0.7713
	MCAR	0.9229	0.0547	78	1.2028
	MNAR	0.8371	0.0582	78.5	1.1725
IV alldichmean	no NAs	0.524	0.049	89.5	0.8677
	MCAR	0.8589	0.0681	86.5	1.2882
	MNAR	0.7379	0.0634	85.5	1.1597
IV alldichmedian	no NAs	0.5228	0.0483	89.5	0.8583
	MCAR	0.874	0.0676	87	1.2934
	MNAR	0.7461	0.0652	84	1.1842
IV prevpatient	no NAs	0.0733	0.2228	95	3.1432
	MCAR	0.0188	0.0873	96	1.2324
	MNAR	-0.0135	0.0529	96	0.7461
IV prev2patient	no NAs	0.0710	0.2268	95.5	3.2
	MCAR	-0.088	0.1119	95.	1.5803
	MNAR	-0.017	0.0595	94.5	0.8392
IV prev5patient	no NAs	0.1868	0.3076	96.5	4.3437
	MCAR	0.1625	0.0952	96.5	1.3533
	MNAR	-0.0259	0.0513	92.5	0.7237
IV prev10patient	no NAs	0.2317	0.4096	96	5.7833
	MCAR	-0.1494	0.1164	96.5	1.6494
	MNAR	0.0448	0.0558	93.5	0.7888

Method	Missing mechanism	Bias	SE	Coverage	RMSE
IV allprevprop	no NAs	-0.0751	0.2498	98	3.525
	MCAR	0.2378	0.158	96.5	2.2412
	MNAR	0.0382	0.0677	94	0.9563

Table 4.14: Summary of performance measures for scenario 2. For this simulation the treatment decision X *is generated using the extended Ertefaie model.*

Appendix 4.5 Simulation study on the classification performance of the Abrahamowicz method

The Abrahamowicz method aims to determine which provider change their prescription preferences and tries to identify the change time i^* for those provider. Here we present the results of a Monte Carlo simulation study to evaluate the performance of this approach. The simulation was conducted using R studio (version 4.2.1). The data generation process is explained in Section 4.4. In 200 simulation runs, study populations of J = 200 providers with $n_j = 10, 50, 100, 500,$ 1000 patients are generated. The treatment decision is simulated using model (4.6). Change in PP is simulated either as abrupt change or as smooth change, explained in Section 4.4.1 or by Abrahamowicz et al. [93].

The classification performance of the algorithm is assessed with commonly used measures calculated from the Confusion matrix. This matrix summarizes the numbers of providers simulated to change PP versus the number of providers that have been identified to change PP by the algorithm to change PP and is given in Table 4.15.

		Identified	
		Change	**No change**
Simulated	**Change**	True positives (TP)	False negatives (FN)
	No change	False positives (FP)	True negatives (TN)

Table 4.15: Confusion matrix contrasting the number of providers simulated with and without a change in PP versus the number of provider that have been identified by the algorithm.

A detailed explanation of the Confusion matrix for binary classifiers and the performance measures derived from this matrix can be found for example in Sokolova and Lapalme [238] and a short summary is given here in Table 4.16.

Performance measures	Formula
Accuracy	$\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FN}+\text{FP}+\text{TN}}$
True positive rate (TPR)	$\frac{\text{TP}}{\text{TP}+\text{FN}}$
True negative rate (TNR)	$\frac{\text{TN}}{\text{TN}+\text{FP}}$
False positive rate (FPR)	$\frac{\text{FP}}{\text{FP}+\text{TN}}$
False negative rate (FNR)	$\frac{\text{FN}}{\text{FN}+\text{TP}}$
Positive predictive value (PPV)	$\frac{\text{TP}}{\text{TP}+\text{FP}}$
Negative predictive value (NPV)	$\frac{\text{TN}}{\text{TN}+\text{FN}}$

Table 4.16: Summary of the performance measures used to assess the classification performance of the Abrahamowicz method.

The simulation results for abrupt change are given in Figure 4.15. Here, the performance measures are summarized in % and as average over all simulation runs for different numbers of patients treated by each provider.

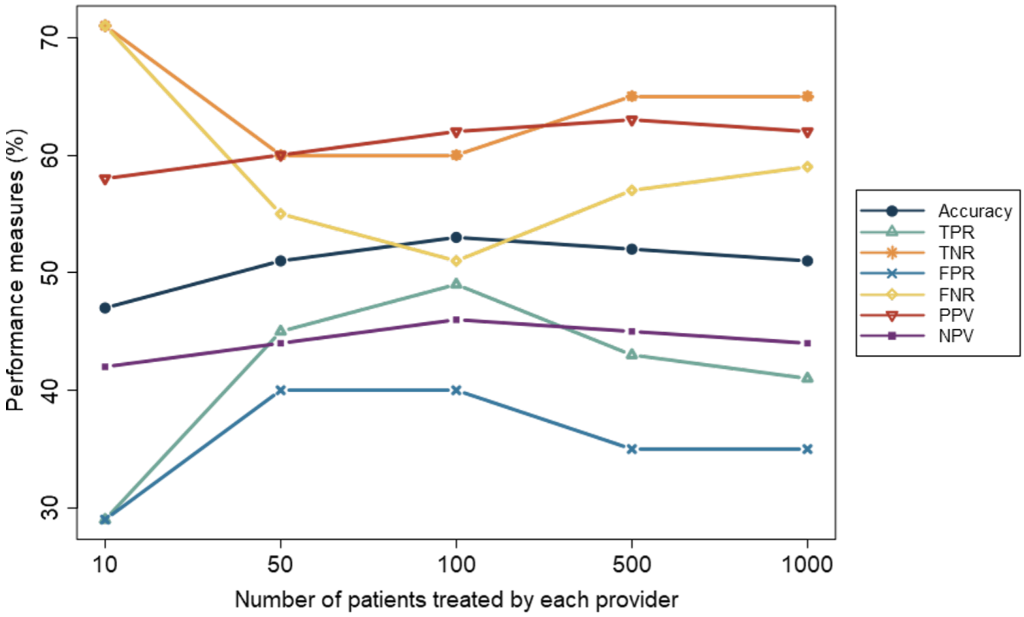


Figure 4.15: Assessment of the classification performance of the Abrahamowicz method for a simulated abrupt change.

The overall effectiveness of the classifier approach is measured with the accuracy. For the abrupt change and over the different sample sizes n_j the accuracy is on average around 50%. From all provider simulated to have a change in $PP_{ji} = 40\%$ are on average identified by the algorithm as indicated with the TPR.

The TNR states that on average 67% of the provider who are simulated not to have a change in PP are identified. These results appear to be consistent over different provider sizes. The PPV reflects on the probability that a provider which has been identified to change PP also had a change simulated and is on average around 60% different provider sizes. Equivalently, the NPV is on average around 45%.

For the simulation employing a smooth change in PP as explain in Section 4.4, the results are visualized in Figure 4.16.

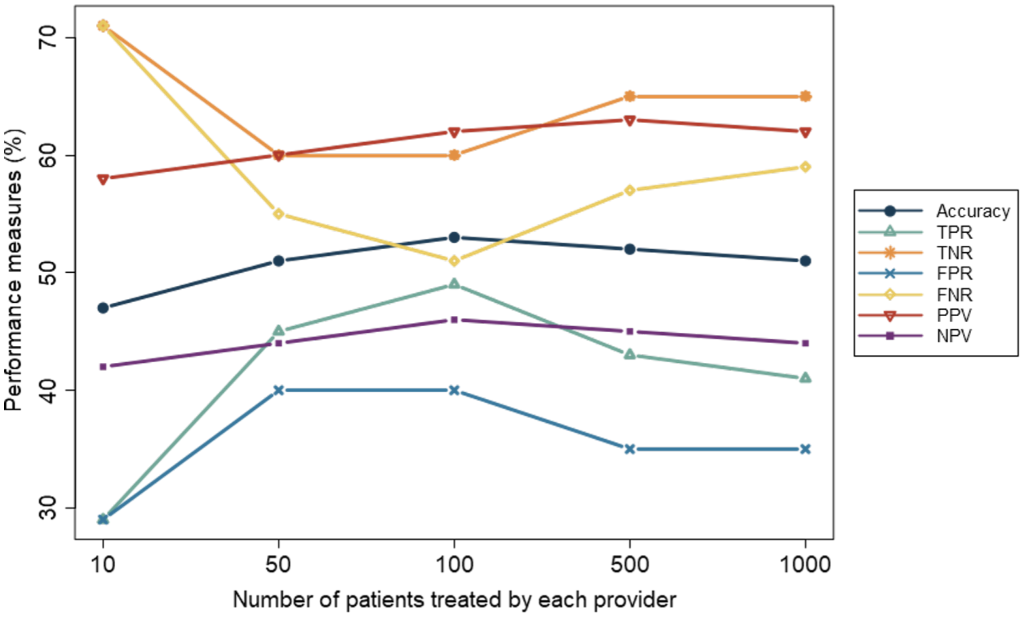


Figure 4.16: Assessment of the classification performance Abrahamowicz method for a simulated smooth change.

They are consistent with the results of the abrupt change simulation giving confidence that the algorithm does not perform worse if change takes place over a longer period. On average the accuracy is around 50% with a TPR of 45% and a TNR of around 65%. The probability to correctly predict provider with a change and without a change are PPV = 60% and NPV = 45% respectively. The results stay also consistent over different provider sizes.

Table 4.17 summarized the mean absolute difference (MAD(i^*)) between simulated and identified change time i^* over all simulation runs, divided by n_j and within the group of providers correctly identified as changing preference (true positives).

		n_j				
		10	50	100	500	1000
$\frac{\text{MAD}(i^\star)}{n_j}$	Abrupt change	0.17	0.26	0.29	0.44	0.47
	Smooth change	0.24	0.32	0.37	0.5	0.5

Table 4.17: Mean absolute difference of simulated and identified $i^\star$*, for the true positive cases divided by* n_j.

The results show that relative to the size of the provider n_j the precision with which $i^\star$ is identified is comparable between the two PP change types.

Appendix 4.6 Additional information on the application case study

A summary of the missing data for the application case study explained in Section 4.5 is given in Table 4.18 for the outcome variable and all measured confounders with missing values. For most of the TSLS estimation, complete case datasets are necessary, except for the Ertefaie method and its extension method. Additionally, practices with too little data n_j for the respective Z construction method are excluded from the analysis. Based on the original study population and for each construction method, a separate dataset is therefore prepared to apply to the IV estimation procedure.

Variable	DPP4i	SGLT2i	Overall
Achieved HbA1c (mmol/mol)	34087 (31.01%)	30189 (39.09%)	64276 (34.4%)
HbA1c (mmol/mol)	11559 (10.5%)	10379 (13.4%)	21938 (11.7%)
BMI (kg/m^2)	5689 (5.2%)	3017 (3.9%)	8706 (4.7%)
eGFR (ml/min/1.73m^2)	726 (0.7%)	347 (0.4%)	1073 (0.6%)
ALT (U/L)	7513 (6.9%)	4783 (6.2%)	12296 (6.6%)
Ethnicity	1882 (1.7%)	1598 (2.1%)	3480 (1.9%)
Deprivation	63 (0.1%)	45 (0.1%)	108 (0.1%)
Smoking status	5368 (4.9%)	3174 (4.1%)	8542 (4.6%)

Table 4.18: Summary of missing values in the study population (% of records missing). The summary shows the outcome variable (achieved HbA1c) and all measured confounders with missing values.

Method	Data size (N)	J	$n_{j,min}$	Complete case information
Observational estimate	103611 SGLT2i: 38.79 %	1403	-	complete case on Y and $\mathbf{W}$
IV ePP	122556 SGLT2i: 38.38 %	1407	2	complete case for Y
IV ePP (rirs)	122556 SGLT2i: 38.38 %	1407	2	complete case for Y
IV star	101377 SGLT2i: 38.95 %	1364	5	complete case for Y and $\mathbf{W}$
IV allprop	103598 SGLT2i: 38.79 %	1390	2	complete case for Y and $\mathbf{W}$
IV alldichmean	103598 SGLT2i: 38.79 %	1390	2	complete case for Y and $\mathbf{W}$
IV alldichmedian	103598 SGLT2i: 38.79 %	1390	2	complete case for Y and $\mathbf{W}$
IV prevpatient	102208 SGLT2i: 39.25 %	1390	2	complete case for Y and $\mathbf{W}$
IV prev2patient	100818 SGLT2i: 39.68 %	1373	3	complete case for Y and $\mathbf{W}$
IV prev5patient	96712 SGLT2i: 40.83 %	1358	6	complete case for Y and $\mathbf{W}$
IV prev10patient	89979 SGLT2i: 42.39 %	1326	11	complete case for Y and $\mathbf{W}$
IV allprevprop	102208 SGLT2i: 38.79 %	1390	2	complete case for Y and $\mathbf{W}$

Table 4.19: Summary of the data preparation for the respective construction methods and the IV estimation. The data preparation process results in different study population sizes (N) and number of providers per dataset (J). Additionally, information on the minimum practice size ($n_{j,min}$) needed ton construct Z and information on the complete case dataset construction is given.

Appendix 4.7 Estimation results of the complete case analysis

The estimation results of the application case study outlines in Section 4.5 are represented in Figure 4.9. These results are estimated using specific datasets for each of the construction method for Z that depend on the complete case data and minimum practice size $n_{j,min}$ required for each of the methods. Only IV ePP and IV ePP (rirs) do not require a complete case dataset on the measured confounders. The two methods show significantly different estimation results compared to other IV methods. The analysis was therefore repeated on a complete case dataset with $n_{j,min} > 11$ on which all IV construction methods can be applied. Figure 4.17 summarizes the results of this analysis. It is noticeable that IV ePP and IV ePP (rirs) still lead to significantly smaller relative treatment effect estimates compared to all other IV construction methods.

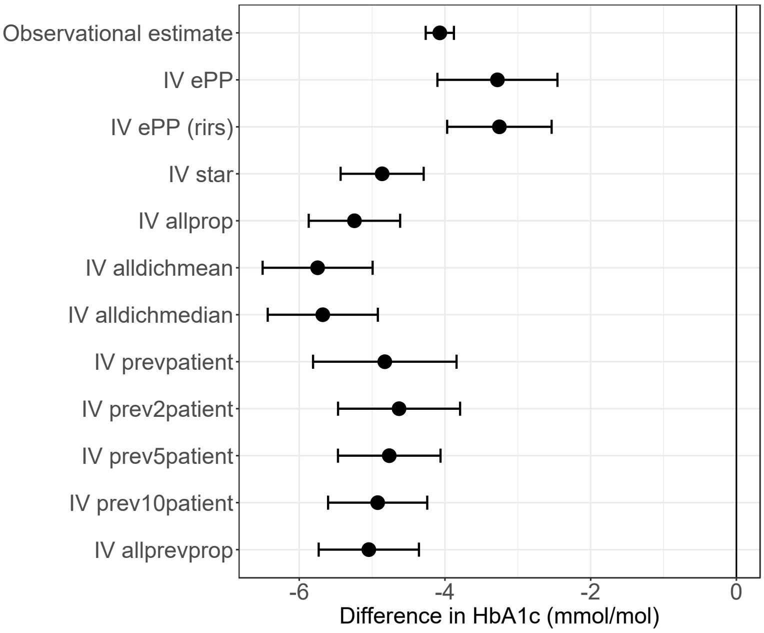


Figure 4.17: Estimation results of the relative treatment effect of SGLT2i versus DPP4i on the reduction of HbA1c (mmol/mol). Values smaller than 0 indicate that SGLT2i has a stronger HbA1c decreasing effect compared to DPP4i. Results are shown for a multivariable regression analysis (observational estimate) and all IV estimates employing the construction methods of a preference-based IV. Estimation procedures was applied on the same complete case dataset with sufficient treatment prescription data for all IV construction methods.

Appendix 4.8 Application case study of Chapter 3 revised

Chapter 3 outlined an application case study evaluating the relative risk of experiencing a genital infection on SGLT2i versus DPP4i for people with T2D initiating the study treatments as second line treatment. We employed a T2D cohort from the CPRD Gold (download July 2019) database. Estimation of the treatment effect of interest for this study was done using conventional estimation methods that account for measured confounders such as multivariable regression model with and without propensity score matched data (CaT and PSM respectively). Additionally, several causal estimation method were employed such as the difference-in-difference approach (DiD) the Instrumental Variable and Control Function approach (IV and CF) as well as the proposed prior outcome augmented Instrumental Variable/ Control Function approach (POA-IV/ POA-CF). As preference-based instrument for this analysis a proxy variable was constructed based on the method proposed by Brookhart et al. [92] (IV prevpatient).

In the following, the analysis is revised and amended using the preference-based instrument construction method by Ertefaie et al. [97]. In the simulation study outlined in this chapter, the Ertefaie method has been proven to be a robust alternative in constructing a proxy instrument in case of non-ignorable missing data and was able to perform well under change in prescription preference over time. These characteristics of the Ertefaie method are of interest for the revised analysis, as the cohort data include missingness for the baseline characteristics: HbA1c (mmol/mol), eGFR ($\mathrm{mL/min/1.73m^2}$), and BMI (kg/m^2). The study outlined in Chapter 3 therefore relied on a complete case analysis.

For the analysis with the Ertefaie method, provider who only treat one patient were excluded. After excluding 17 provider with too little prescription data and 4 provider with no fully recorded patient records regarding baseline characteristics, data from 419 provider was used for the analysis. Average number of patients treated by each provider was 21. The smallest provider treated 2 patients and the largest provider 108 patients. In the simulation study outlined in this chapter, model-based construction methods performed better in case of larger provider sample sizes. Figure 4.18 and Table 4.20 show the results of the CaT, PSM, IV prevpatient, DiD (described in Chapter 3) and the results of the Ertefaie method, IV ePP for consistency comparison of the estimation results in this triangulation framework. Results are shown on risk difference scale derived using the `margins()` package [215] as explained in Chapter 3.

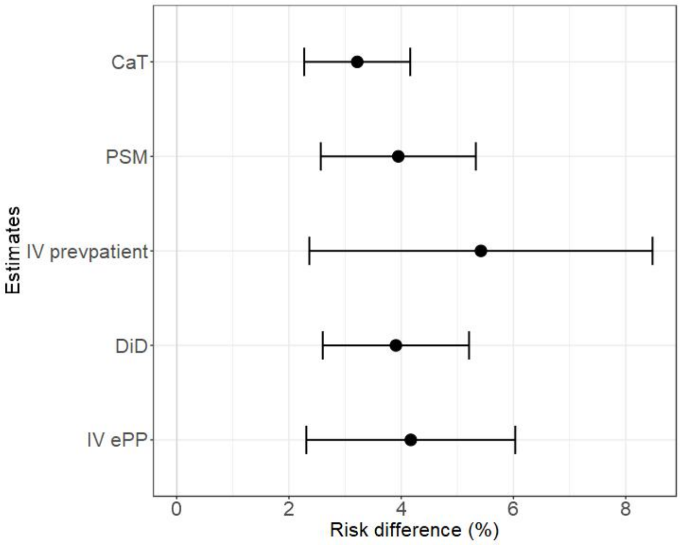


Figure 4.18: Estimation results of the revised application case study outlined in Chapter 3. Results are shown for the of multivariable regression with and without propensity score matched data (CaT and PSM), difference-in-difference approach (DiD), IV approach using the preference-based instrument construction method proposed by Brookhart et al. (IV prevpatient) and the IV approach using the Ertefaie method (IV ePP) shown with their 95% confidence intervals.

Method	Estimate	95% CI	SE	p-value
CaT	3.22	2.27, 4.16	0.49	3.10×10^{-14}
PSM	3.95	2.57, 5.33	0.72	5.72×10^{-10}
IV prevpatient	5.42	2.36, 8.48	2.42	0.0003
DiD	3.91	2.6, 5.21	0.66	7.46×10^{-10}
IV ePP	4.17	2.31, 6.03	0.95	5.06×10^{-6}

Table 4.20: Estimation results on risk difference scale (in %) for the revised application case study of Chapter 3, standard error, and p-value of the estimated treatment effect.

The estimation results are consistent for all methods shown in the figure and conclude that relative risk of experiencing a genital infection is increased on SGLT2is.

The Ertefaie method is able to estimate the treatment effect more efficiently compared to the revised results of IV prevpatient which is in line with simulation findings outlined in this chapter. Furthermore, IV ePP is able to estimate the relative risk with comparable efficiency compared to DiD, CaT and PSM. Table 4.21 summarizes the IV strength of IV prevpatient and IV ePP and show that both instrument lead to an F-statistic greater than 10. This gives confidence for a lack of weak instrument bias in the estimation results.

Models	F-statistic
IV prevpatient	345.42
IV ePP	20.82

Table 4.21: Strength of the instrumental variables measured with the F-statistic of Z *for IV prevpatient and IV ePP from the corresponding first stage regression models and the revised application case study of Chapter 3.*

Chapter 5

Evaluation of the safety and effectiveness of SGLT2 Inhibitors in adults over 70 using an Instrumental Variable approach: UK population based study

Laura M. Güdemann, Katie G. Young, Nicholas J. M. Thomas, Rhian Hopkins, Robert Challen, Angus Jones, Andrew Hattersley, Beverley M. Shields, Jack Bowden, John M. Dennis, & Andrew P. McGovern
on behalf of the MASTERMIND consortium

Planned submission in: Diabetes Care

Author contribution

LMG, APM, JMD, BMS and JB designed the study. APM, NT, AJ and AH provided invaluable clinical insight and helped interpreting the results. KGY, RH, RC and APM developed code lists for the identification of relevant outcomes and comorbidities for the construction of the type 2 diabetes cohort. AC, KGY, RH, JMD and BMS constructed the type 2 diabetes cohort of the CPRD data. From this cohort, LMG identified patients with relevant treatment regimes and characteristics for this study. LMG, JMD, BMS and JB developed the analysis strategy. LMG analysed the data under supervision of JMD, BMS and JB. LMG drafted the original version of the paper which APM, JMD, BMS and JB helped to edit.

5.1 Abstract

Objective

Current type 2 diabetes guidelines recommend an individualised approach to treatment, but lack evidence based guidance for the heterogeneous patient group of the older adults, particularly in real world data. We aimed to develop a causal analysis framework to assess the safety and effectiveness profile of Sodium-glucose Cotransporter-2 Inhibitors (SGLT2i) in this patient group.

Research design and methods

Routine primary care data from the large Clinical Practice Research Datalink UK cohort, linked to hospital records, was used to compare the relative risk profile and relative effectiveness of SGLT2i and Dipeptidyl peptidase-4 Inhibitors (DPP4i). We analysed treatment and adverse effect outcomes in patients initiating SGLT2i and DPP4i between 2013 and 2020. Analysis was stratified by age <70 years (SGLT2i n = 66810, DPP4i n = 76172) and $\geq$70 years (SGLT2i n = 10,419, DPP4i n = 33,434). Study outcomes were assessed using the Instrumental Variable method and a proxy measure of prescription preference as instrument.

Results

There was no evidence of an increased risk of volume depletion, poor micturition control, urinary frequency, falls or amputation in people over or under 70 years initiating SGLT2i compared to DPP4i. Risk of diabetic ketoacidosis (DKA) was increased with SGLT2i in those $\geq$70 years (relative risk (RR) 3.82 [CI 95% 1.12, 13.03]), but was not observed in those <70 years (RR 1.12 [CI 95% 0.41, 3.04]). SGLT2i were associated with a similarly increased risk of genital infection in both age groups (RR 2.27 [CI 95% 2.03, 2.53] for <70 years; RR years 2.16 [RR CI 95% 1.77, 2.63] for $\geq$70). In those $\geq$70 years, HbA1c reduction was similar on both SGLT2i and DPP4i (HbA1c benefit with SGLT2i: -0.3 mmol/mol [CI 95% -1.6, 1.1]), but was greater with SGLT2i in those <70 years (-4 mmol/mol [CI 95% -4.8, -3.1]). Weight reduction was consistently greater with SGLT2i compared to DPP4i in both age groups.

Conclusions

Causal analysis using large-scale observational data suggests SGLT2is are effective and generally safe in older adults, but increase risk of both genital infections and, rarely, DKA. By extending evidence from RCTs to understand the risk and benefit profile of SGLT2i to older adults with type 2 diabetes, our study supports careful prescribing of SGLT2i in this important patient population.

5.2 Introduction

Current type 2 diabetes guidelines recommend an individualised approach to treatment that takes into account patient preferences, comorbidities, risks from polypharmacy, and the likelihood of benefiting from long-term interventions, [103, 239] but clear guidance on therapeutic strategies for the management of type 2 diabetes in older patients, for example over the age of 70 years, is limited. [113] For the older patients, specific treatment considerations are likely to be needed, due to increased comorbidities, age-related changes in physiology and pharmacodynamics, as well as possible increased propensity to adverse medication effects. Additionally, due to limited life expectancy, long-term glycaemic control benefits are less relevant than in younger patients. [133, 140]

Under current guidelines, a large proportion of older type 2 diabetes patients would be recommended SGLT2i due to their cardiorenal benefits, and irrespective of patients' glycaemic control. [103] SGLT2is have well described benefits, particularly cardiorenal and the promotion of weight loss [161, 182, 240, 241, 242], but also possible risks, which may limit their use for older patients. [113] Well-established risks of SGLT2is are genital infections and due to their mode of action, volume depletion is possible. [164, 182] These side effects could be of particular concern for the older adults where incontinence, dehydration and dizziness could have more severe consequences compared with a younger population. [133, 165, 166, 167] Additionally, dehydration or dizziness can also lead to falls in older adults. [154] Further adverse effects of concerns of SGLT2is are lower limb amputations [164]. Reports of possible association of SGLT2i and DKA as prompted the FDA [189] and the EMA [190] to issue warnings. DKA is of special concern for older type 2 diabetes patients as they often have long-duration diabetes causing low residual insulin secretion. Older people may also present with more frequent acute complications, such as infections, which are additional risk factors of DKA. [162]

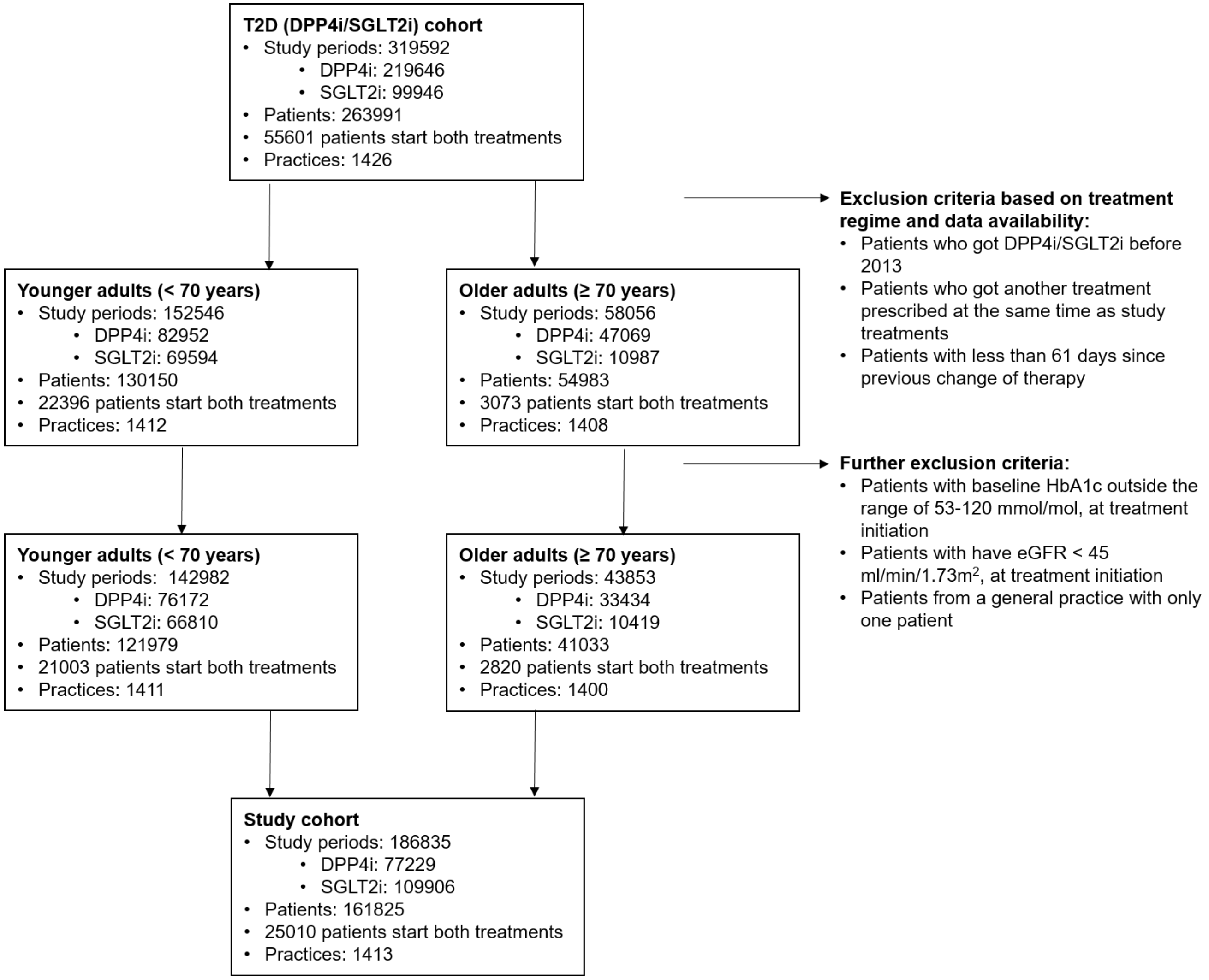
In order to develop targeted guidelines for the management of type 2 diabetes in older adults, strong evidence-based data are needed. [113] However, guidelines on glycaemic targets, therapeutic interventions and descriptions of risk profiles are informed by randomized clinical trials (RCTs), which often exclude individuals over 65 years due to common comorbidities. This means caution is needed when extrapolating RCT evidence for this patient group. [113, 140, 143] Most RCTs are not designed for the older patient population and do not take functional status into account, so do not represent a real-world population of older type 2 diabetes patients. [189] Observational studies of older type 2 diabetes patient population are rare but have the potential to provide insights that cannot necessarily be given by RCTs. Previous post-hoc RCT analyses [165, 175, 176, 243] have examined risks in older patients, but have very small sample sizes for this patient subgroup, and therefore might suffer from potential outlier effects. [165] Also, without detailed data on patient characteristics, comorbidities and concomitant medications the results from observational studies may be affected by unmeasured confounding which can bias treatment effect results. [154]

Given the lack of robust studies, we aimed to examine the relative risks and benefits of SGLT2i in older patients, over 70 years, compared to the most commonly prescribed second-line diabetes drug class, DPP4i using routine primary care data. We employ an Instrumental Variable approach, exploiting systematic variation in practitioners' prescribing preference as the instrument, to estimate the impact of receiving SGLT2i compared to DPP4i on a range of adverse effects and important treatment outcomes, analogous to a randomised controlled trial.

5.3 Methods

5.3.1 Study design and participants

In this retrospective cohort study, the UK routine primary care data were accessed from Clinical Research Datalink (CPRD) Aurum (November 2020 download). CPRD is a UK representative sample covering approximately 6.9% of the population. [200] CPRD Aurum was linked to Hospital Episode Statistics (HES), Office for National Statistics (ONS) death registrations and patient-level Index of Multiple Deprivation (IMD). Type 2 diabetes patients were identified according to a previously published protocol [203] based on the presence of a diagnostic code for diabetes and the prescription of one or more glucose lowering medications. Type 1 diabetes and other types of diabetes were excluded. The analysis included new users of SGLT2i (Canagliflozin, Dapagliflozin, Empagliflozin, Ertugliflozin), initiating treatment after 1st January 2013 and with an identifiable date of type 2 diabetes diagnosis. The comparison cohort was new users of DPP4i (Alogliptin, Linagliptin, Sitagliptin, Saxagliptin, Vildagliptin), as these agents represent the most commonly prescribed drug class after metformin in the UK, and have no known association with the SGLT2i-associated adverse effects of interest evaluated in this study. All available follow-up data was considered in the analysis up to the point of data extraction. Patients with a baseline HbA1c outside of the range 53-120 mmol/mol were excluded from the analysis, reflecting on the threshold for glucose-lowering medication initiation in clinical guidelines and severe hyperglycemia. Additionally, patients with renal impairment indicated with a glomerular filtration rate (eGFR) of less than 45 mL/min/1.73 m^2 were excluded, as SGLT2i was not licensed for use below this threshold for the majority of the study period. Further exclusion criteria are summarized in Figure 5.1. Our cohort was split into a younger (<70 years at treatment initiation) and older ($\geq$70 years) population.



T2D (DPP4i/SGLT2i) cohort
- Study periods: 319592
    - DPP4i: 219646
    - SGLT2i: 99946
- Patients: 263991
- 55601 patients start both treatments
- Practices: 1426

Exclusion criteria based on treatment regime and data availability:
- Patients who got DPP4i/SGLT2i before 2013
- Patients who got another treatment prescribed at the same time as study treatments
- Patients with less than 61 days since previous change of therapy

Younger adults (< 70 years)
- Study periods: 152546
    - DPP4i: 82952
    - SGLT2i: 69594
- Patients: 130150
- 22396 patients start both treatments
- Practices: 1412

Older adults (≥ 70 years)
- Study periods: 58056
    - DPP4i: 47069
    - SGLT2i: 10987
- Patients: 54983
- 3073 patients start both treatments
- Practices: 1408

Further exclusion criteria:
- Patients with baseline HbA1c outside the range of 53-120 mmol/mol, at treatment initiation
- Patients with have eGFR < 45 ml/min/1.73m^2, at treatment initiation
- Patients from a general practice with only one patient

Younger adults (< 70 years)
- Study periods: 142982
    - DPP4i: 76172
    - SGLT2i: 66810
- Patients: 121979
- 21003 patients start both treatments
- Practices: 1411

Older adults (≥ 70 years)
- Study periods: 43853
    - DPP4i: 33434
    - SGLT2i: 10419
- Patients: 41033
- 2820 patients start both treatments
- Practices: 1400

Study cohort
- Study periods: 186835
    - DPP4i: 77229
    - SGLT2i: 109906
- Patients: 161825
- 25010 patients start both treatments
- Practices: 1413

Figure 5.1: Flow chart of study cohort selection, age stratified

5.3.2 Outcomes

Adverse effects (AE) included in analysis were genital infections, micturition control, volume depletion and dehydration, urinary frequency, falls, lower limb amputation and diabetic ketoacidosis. An occurrence of each AE was measured up to 3 years after treatment initiation and censoring of the follow-up time was implemented in case of a discontinuation of the study treatment or start of the comparison study treatment. Individuals were therefore followed up until the earliest of: date of the outcome of interest, discontinuation of the study treatment, start of comparison study treatment, date of practice deregistration/death, end of study period, or 3 years. Occurrences of AEs were identified using diagnosis code lists published at: `https://github.com/Exeter-Diabetes/CPRD-Codelists`. Genital infections were identified with either a diagnosis code for a specific genital infection (e.g. candida vaginitis or vulvo-vaginitis in women, balanitis, balanoposthitis in men), a prescription for antifungal therapy used specifically to treat genital infections (e.g. an antifungal vaginal pessary), or a non-specific diagnosis of 'thrush' with a topical antifungal prescribed on the same day. [157] The diagnosis codes to identify amputation AEs were taken from Pearson-Stuttard et al. [244]. DKA was identified using HES hospitalization data. Treatment outcomes to assess relative effectiveness of SGLT2i included achieved glycated haemoglobin (HbA1c in mmol/mol) and weight (kg). These outcome measurements were taken as the closest recorded value to 12 months post treatment initiation, within a window of 3 to 15 months.

5.3.3 Covariates

Measured covariates for all outcome models were extracted following our previous protocol [203] and included general information about patients, such as sociodemographic features (age, sex, ethnicity and deprivation) and treatment history, important biomarkers as well as history of relevant comorbidities. Biomarker baseline values are defined nearest to treatment initiation up to 2 years before and 7 days after initiation. Initiation of relevant additional treatments such as diuretics, have been observed up to 3 months before treatment initiation and comorbidities have been characterizes to be within 1 year, 1-5 years or >5 to treatment initiation. A summary of all covariates is given in Table 5.1, a cohort description and a comprehensive overview of the biomarker and comorbidity definitions are given here: `https://github.com/Exeter-Diabetes/CPRD-Cohort-scripts`.

5.3.4 Statistical methods

Unadjusted analysis

In order to contrast causal effect results, absolute risk results for AEs were calculated form survival models without adjusting for measured baseline characteristics. Furthermore, mean achieved treatment outcomes for both treatment and age groups were calculated as the sample average of achieved HbA1c (mmol/mol) and weight (kg).

Causal analysis

When analysing treatment effects from observational data, bias due to confounding by indication is a major challenge. The confounding pre-treatment variables affect the outcome and the treatment decision simultaneously. As a result, it is possible that they differ in distribution between patient who received the study and comparator treatment. [245] Traditional methods such as propensity score matching can mitigate the risk of bias by adjusting for measured confounders, but they cannot control for variables that are not recorded in the data which can lead to unmeasured confounding. [245] With the Instrumental Variable (IV) approach and given a suitable instrument, treatment effects can be estimated in the presence of residual or unmeasured confounding without bias. [246] The basic idea of the IV approach is that a suitable IV is used to extract variation of the treatment that is free of unmeasured confounding. This variation is then utilized to estimate the treatment effect. [245] We employ the IV approach proposed by Ertefaie et al. [97] which makes use of observed treatment behaviour and covariates to construct a proxy for prescription preference. Additionally, the method is capable of estimating the treatment effect without bias even in the presence of non-ignorable missingness in covariates. Our analysis did therefore not rely on a possibly selective complete case dataset. A more detailed explanation of this approach and a description of the assumed data structure for this study can be found in the Appendix 5.1.

All binary AE outcomes were modelled using generalized Poisson regression with follow-up time (in days) as offset. For the estimation of the treatment effect of SGLT2i on achieved HbA1c and weight a linear outcome model was used. Models used in the IV estimation and for all outcomes of interest were adjusted using slightly different sets of relevant covariates. A summary of all models is provided in the Appendix 5.2.

Additionally, conventional multivariable regression models utilizing unmatched and propensity score matched data as well as the Instrumental Variable method based on the preference-based instrument constructed with the previous prescription for each patient are applied to estimate the treatment effects. Comparing the estimation results of these methods will make it possible to judge the robustness of the IV method applied in the main analysis. Further details on this triangulation analysis and discussion of the results are provided in Appendix 5.6.

All analyses outlined in this chapter has been conducted in R Studio (version 4.2.1), online supplementary material including R codes for the estimation of the outcome models is available here: `https://github.com/Exeter-Diabetes/CPRD-Laura-SGLT2i-in-older-adults`.

5.3.5 Sensitivity analysis

As incidence rates were low for volume depletion/ dehydration, micturition control and urinary frequency, we derived a composite outcome for osmotic symptoms for the causal analysis. We also constructed a composite outcome for falls and lower limb fracture. Not all falls might be coded in the CPRD data and lower limb fractures are often caused by falls. Our code list for lower limb fractures excludes fractures of the foot but includes hip fractures which are caused 98% of the times by a fall. [247] As additional sensitivity analysis, censoring was applied in any case of treatment regime change, meaning we additionally censored patients who switched or added any other type 2 diabetes treatments other than the study treatments. We also conducted an analysis using 1 year maximum follow-up time for AE outcomes to assess short term risks. The analysis was done additionally, on a cohort which excluded the second study period of all patients initiating both study treatments. For the main analysis, these patients contributed with both respective study periods.

5.4 Results

Figure 5.1 represents a flow chart of the study cohort selection and the applied exclusion criteria. The study cohort included 186835 study periods from 161825 patients (25010 patients initiated both treatments). There were 142982 study periods included in the analysis for adults under 70 years (n = 76172 SGLT2i, n = 66810 DPP4i) and 43853 study periods for adults 70 years and older (n = 10419 SGLT2i, n = 33434 DPP4i). Table 5.1 shows the baseline characteristics of the study population by treatment arm and age group. In the Appendix 5.3 a more detailed summary of comorbidity history is provided as well as a summary of the amount of missing data in the respective covariates. Furthermore, the person-years and average follow-up time (in years) are also summarized for each AE outcome in the Appendix 5.3.

	SGLT2i	SGLT2i	DPP4i	DPP4i
	< 70 years	$\geq$ 70 years	< 70 years	$\geq$ 70 years
	(n = 66810)	(n = 10419)	(n = 76172)	(n = 33434)
Age (years)	55.8 (8.83)	74.5 (3.81)	56.7 (8.98)	77.3 (5.37)
Sex				
Male	40863 (61.2)	6344 (60.9)	47185 (61.9)	18449 (55.2)
Female	25947 (38.8)	4075 (39.1)	28987 (38.1)	14985 (44.8)
HbA1c (mmol/mol)	77.6 (15.0)	74.8 (13.8)	74.1 (14.5)	71.0 (12.9)
eGFR $(\mathrm{ml/min/1.73m^2})$	97.1 (14.3)	80.4 (12.5)	94.1 (16.4)	73.1 (15.4)
ALT (U/L)	35.6 (20.5)	27.6 (15.2)	34.8 (20.5)	24.9 (14.6)
BMI $(\mathrm{kg/m^2})$	34.2 (6.9)	31.6 (5.8)	32.7 (6.8)	30.0 (5.6)
Weight (kg)	98.9 (22.1)	89.2 (18.3)	94.1 (21.4)	83.3 (17.5)
Insulin ever taken				
Yes	57484 (86)	9054 (86.9)	72872 (95.7)	31423 (94)
No	9326 (14)	1365 (13.1)	3300 (4.3)	2011 (6)
T2D duration (years)	9.33 (6.07)	13.2 (6.99)	7.77 (5.7)	11.8 (7.4)
Alogliptin			15088 (19.8)	6901 (20.6)
Linagliptin			14657 (19.2)	10820 (32.3)
Saxagliptin			4507 (5.9)	1725 (5.2)

	SGLT2i < 70 years (n = 66810)	SGLT2i $\geq$ 70 years (n = 10419)	DPP4i < 70 years (n = 76172)	DPP4i $\geq$ 70 years (n = 33434)
Sitagliptin			41281 (54.2)	13717 (41)
Vildagliptin			639 (0.8)	271 (0.8)
Canagliflozin	11307 (16.9)	2177 (20.9)		
Dapagliflozin	30253 (45.3)	3701 (35.5)		
Empagliflozin	25181 (37.7)	4524 (43.4)		
Ertugliflozin	69 (0.1)	17 (0.2)		
Number of concurrent treatments				
1	3554 (5.3)	739 (7.1)	5877 (7.7)	5375 (16.1)
2	29891 (44.7)	3892 (37.4)	45043 (59.1)	18475 (55.3)
3+	33365 (49.9)	5788 (55.6)	25252 (33.2)	9584 (28.7)
Drugline				
1	523 (0.8)	48 (0.5)	1404 (1.8)	1057 (3.2)
2	13346 (20)	1282 (12.3)	32001 (42)	11886 (35.6)
3	18475 (27.7)	2566 (24.6)	30650 (40.2)	13847 (41.4)
4+	34466 (51.6)	6523 (62.6)	12117 (15.9)	6644 (19.9)
Year of treatment initiation				
2013	1127 (1.7)	127 (1.2)	9305 (12.2)	3345 (10)
2014	4971 (7.4)	566 (5.4)	9499 (12.5)	3539 (10.6)
2015	8910 (13.3)	1245 (11.9)	10542 (13.8)	4290 (12.8)
2016	9805 (14.7)	1316 (12.6)	11745 (15.4)	4959 (14.8)
2017	10904 (16.3)	1494 (14.3)	11659 (15.3)	5300 (15.9)
2018	12271 (18.4)	2054 (19.7)	11016 (14.5)	5310 (15.9)
2019	13320 (19.9)	2542 (24.4)	9059 (11.9)	4910 (14.7)
2020	5502 (8.2)	1075 (10.3)	3347 (4.4)	1781 (5.3)
Ethnicity				
White	50321 (75.3)	9072 (87.1)	55279 (72.6)	28787 (86.1)
South Asian	10172 (15.2)	791 (7.6)	12576 (16.5)	2450 (7.3)
Black	3086 (4.6)	266 (2.6)	4580 (6)	1342 (4)
Other	1041 (1.6)	107 (1)	1348 (1.8)	299 (0.9)
Mixed	722 (1.1)	53 (0.5)	863 (1.1)	200 (0.6)

	SGLT2i < 70 years (n = 66810)	SGLT2i $\geq$ 70 years (n = 10419)	DPP4i < 70 years (n = 76172)	DPP4i $\geq$ 70 years (n = 33434)
Deprivation index				
1	5127 (7.7)	1169 (11.2)	5150 (6.8)	3516 (10.5)
2	5476 (8.2)	1169 (11.2)	5622 (7.4)	3602 (10.8)
3	5673 (8.5)	1134 (10.9)	6212 (8.2)	3633 (10.9)
4	5707 (8.5)	1140 (10.9)	6410 (8.4)	3368 (10.1)
5	6101 (9.1)	1015 (9.7)	6581 (8.6)	3402 (10.2)
6	6679 (10)	1018 (9.8)	7616 (10)	3407 (10.2)
7	7581 (11.3)	1078 (10.3)	8645 (11.3)	3539 (10.6)
8	7691 (11.5)	925 (8.9)	9313 (12.2)	3083 (9.2)
9	8425 (12.6)	908 (8.7)	10570 (13.9)	3067 (9.2)
10	8311 (12.4)	857 (8.2)	10013 (13.1)	2794 (8.4)
Smoking status				
Active smoker	11793 (17.7)	951 (9.1)	14803 (19.4)	2968 (8.9)
Ex-smoker	35054 (52.5)	6806 (65.3)	37892 (49.7)	20718 (62)
Non-smoker	17275 (25.9)	2176 (20.9)	19927 (26.2)	7930 (23.7)
Loop diuretics	2428 (3.6)	997 (9.6)	3288 (4.3)	4836 (14.5)
Ksparing diuretics	1185 (1.8)	314 (3)	1507 (2)	1298 (3.9)
Thiazide diuretics	7730 (11.6)	1772 (17)	9312 (12.2)	5916 (17.7)
Immunosuppressants	625 (0.9)	144 (1.4)	838 (1.1)	428 (1.3)
Oestrogens	853 (1.3)	69 (0.7)	950 (1.2)	314 (0.9)
Oral steroids	1579 (2.4)	454 (4.4)	2274 (3)	1993 (6)
Statins	48595 (72.7)	8132 (78)	54851 (72)	25313 (75.7)
ACE inhibitors	28655 (42.9)	4714 (45.2)	31242 (41)	14529 (43.5)
History genital infection				
Yes	34577 (51.8)	5277 (50.6)	36903 (48.4)	16432 (49.1)
No	32233 (48.2)	5142 (49.4)	39269 (51.6)	17002 (50.9)
History urinary frequency				
Yes	6530 (9.8)	1638 (15.7)	7499 (9.8)	5365 (16)
No	60280 (90.2)	8781 (84.3)	68673 (90.2)	28069 (84)
History micturition control				
Yes	6002 (9)	1247 (12)	6866 (9)	5059 (15.1)
No	60808 (91)	9172 (88)	69306 (91)	28375 (84.9)
History volume depletion				
Yes	5630 (8.4)	1147 (11)	6369 (8.4)	4548 (13.6)
No	61180 (91.6)	9272 (89)	69803 (91.6)	28886 (86.4)

	SGLT2i < 70 years (n = 66810)	SGLT2i ≥ 70 years (n = 10419)	DPP4i < 70 years (n = 76172)	DPP4i ≥ 70 years (n = 33434)
History benign prostatehyperplasia				
Yes	2200 (3.3)	1448 (13.9)	2963 (3.9)	5057 (15.1)
No	64610 (96.7)	8971 (86.1)	73209 (96.1)	28377 (84.9)
History lower limb fractures				
Yes	4650 (7)	851 (8.2)	4948 (6.5)	3061 (9.2)
No	62160 (93)	9568 (91.8)	71224 (93.5)	30373 (90.8)
History falls				
Yes	7907 (11.8)	2376 (22.8)	8921 (11.7)	9300 (27.8)
No	58903 (88.2)	8043 (77.2)	67251 (88.3)	24134 (72.2)
History amputation				
Yes	333 (0.5)	51 (0.5)	415 (0.5)	282 (0.8)
No	66477 (99.5)	10368 (99.5)	75757 (99.5)	33152 (99.2)
History diabetic ketoacidosis				
Yes	431 (0.6)	31 (0.3)	367 (0.5)	166 (0.5)
No	66379 (99.4)	10388 (99.7)	75805 (99.5)	33268 (99.5)
History dementia				
Yes	153 (0.2)	189 (1.8)	274 (0.4)	1674 (5)
No	66657 (99.8)	10230 (98.2)	75898 (99.6)	31760 (95)
History cancer				
Yes	3833 (5.7)	1653 (15.9)	5160 (6.8)	6415 (19.2)
No	62977 (94.3)	8766 (84.1)	71012 (93.2)	27019 (80.8)
History asthma				
Yes	13678 (20.5)	1962 (18.8)	14372 (18.9)	6247 (18.7)
No	53132 (79.5)	8457 (81.2)	61800 (81.1)	27187 (81.3)
History COPD				
Yes	3684 (5.5)	1223 (11.7)	4692 (6.2)	4411 (13.2)
No	63126 (94.5)	9196 (88.3)	71480 (93.8)	29023 (86.8)
History heart failure				
Yes	2437 (3.6)	907 (8.7)	3169 (4.2)	4141 (12.4)
No	64373 (96.4)	9512 (91.3)	73003 (95.8)	29293 (87.6)
History CVD				
Yes	13131 (19.7)	3841 (36.9)	15349 (20.2)	14067 (42.1)
No	53679 (80.3)	6578 (63.1)	60823 (79.8)	19367 (57.9)

	SGLT2i < 70 years (n = 66810)	SGLT2i ≥ 70 years (n = 10419)	DPP4i < 70 years (n = 76172)	DPP4i ≥ 70 years (n = 33434)
History CLD				
Yes	8366 (12.5)	959 (9.2)	8093 (10.6)	2243 (6.7)
No	58444 (87.5)	9460 (90.8)	68079 (89.4)	31191 (93.3)
History osteoporosis				
Yes	666 (1)	384 (3.7)	75248 (98.8)	31514 (94.3)
No	66144 (99)	10035 (96.3)	924 (1.2)	1920 (5.7)

Table 5.1: Baseline characteristics of the study cohort. Values for continuous variables are given in mean (standard deviation) and for binary and categorical variables in n (%). COPD: chronic obstructive pulmonary disease, CVD: composite of myocardial infarction, stoke, revascularisation, ischemic heart disease, angina, peripheral arterial disease, transient ischemic attack, CLD: chronic liver disease.

Causal analysis showed increased risk of genital infection for patients initiating SGLT2i but no differences between the age groups

Results of the relative risk estimate for the AEs of interest are summarized in Figure 5.2. Genital infections were most often recorded. In younger adults, there were 12596 events and incidence rates (IR) per 1000 person-years of 102.96 [CI 95% 102.9, 103.02] in those initiating DPP4i and 18493 events (IR 195.81 [CI 95% 195.72, 195.9]) for patients initiating SGLT2i. In older adults, 5406 (IR 95.69 [CI 95% 95.61, 95.77]) events for patients on DPP4i and 2655 (IR 195.35 [CI 95% 195.12, 195.59]) for patients on SGLT2i were recorded. Our causal analysis shows that initiation of SGLT2i significantly increase the relative risk for genital infection, but was similar in both patient groups. The relative risk of experiencing a genital infection with SGLT2i compared to DPP4i was 2.27 [CI 95% 2.03, 2.53] for the younger adults versus 2.16 [CI 95% 1.77, 2.63] for older adults.

DKA was a rare AE and our causal analysis showed increased risk for older patients

Only very few DKA events were recorded in the study cohort. 125 events (IR 0.9 [CI 95% 0.89, 0.91]) were recorded for those younger adults initiating DPP4i and 255 events (IR 2.13 [CI 95% 2.12, 2.13]) for patients on SGLT2i. For older adults on DPP4i 112 events (IR 1.75 [CI 95% 1.75, 1.76]) were recorded and 50 events (IR 2.96 [CI 95% 2.93, 2.99]) for patients on SGLT2i. Relative risk results for DKA was 3.82 [CI 95% 1.12, 13.03] times higher for patients initiating SGLT2i compared to DPP4i in the patient group of older adults. No significant difference was found for the younger patient group with RR 1.12 [CI 95% 0.41, 3.04] between patients on SGLT2i or DPP4i. This is on a background of a higher DKA rate in older adults with type 2 diabetes.

No increased relative risk for patients initiating SGT2i for falls or amputations was found for both age groups

In the group of younger patients 2992 events of falls (IR 22.08 [CI 95% 22.05, 22.1]) were recorded for patients initiating DPP4i and 2253 events (IR 19.17 [CI 95% 19.15, 19.2]) for patients initiating SGLT2i. For older patients initiating DPP4i 6093 events (IR 106.53 [CI 95% 106.53, 106.7]) versus 1016 events (IR 64 [CI 95% 63.87, 64.12]) initiating SGLT2i were recorded. From our causal analysis no relative risk increase of falls for both age groups can be concluded with RR 0.86 [CI 95% 0.66, 1.13] for younger adults and RR 0.56 [CI 95% 0.45, 0.7] for older adults. Also, no significant increase in relative risk was concluded for lower limb amputations with RR 0.58 [CI 95% 0.22, 1.53] for younger adults and RR 1.14 [CI 95% 0.29, 4.57] for older adults. Lower limb amputation was a rare AE in this study cohort and therefore, statistical power to test this effect might have been low. In the younger patient group 192 events (IR 1.38 [CI 95% 1.38, 1.39]) for DPP4i and 199 events (IR 1.66 [CI 95% 1.65, 1.67]) were recorded. In the older patient group 117 events (IR 1.83 [CI 95% 1.82, 1.84]) on DPP4i and 34 events (IR 2.02 [CI 95% 1.99, 2.04]) on SGLT2i were recorded.

Higher relative efficacy was found for SGLT2i, except for glycaemic response in the older patient group

Causal estimate results for HbA1c response and weight change are shown in Figure 5.3. For younger adults, there was on average a -4 mmol/mol [CI 95% -4.8, -3.1] greater reduction in HbA1c with SGLT2i compared to DPP4i. For older adults, HbA1c response on both drug classes was similar with an average relative response of -0.25 mmol/mol [CI 95% -1.63, 1.13]. In contrast, there was a greater reduction in weight with SGLT2i compared to DPP4i in both age groups, but no differences between them. Weight difference for younger adults was on average -2.6 kg [CI 95% -3, -2.3] and -2.8 kg [CI 95% -3.3, -2.3] for older patients.

Unadjusted results showed elevated risk for most AEs and higher effectiveness of SGLT2i

Results of the unadjusted analysis are summarized for the AE and treatment outcomes in the Appendix 5.4. The unadjusted absolute risk analysis overall led to noticeably different conclusions to our causal estimates and showed an increased absolute risk for almost all AEs for both treatments. Different from the causal analysis, the unadjusted absolute risk results showed increased risk for all osmotic symptoms for both treatments and age groups. Unadjusted absolute DKA risk was 0.67 [CI 95% 0.59, 0.76] on SGLT2i and 0.27 [CI 95% 0.22, 0.32] on DPP4i for younger adults and 0.9 [CI 95% 0.63, 1.18] on SGLT2i and 0.53 [CI 95% 0.42, 0.63] on DPP4i for the older study population. Furthermore, unadjusted mean average change results shown a higher reduction in HbA1c (mmol/mol) and weight (kg) for patients initiating SGLT2i, but age-group differences only for HbA1c.

Results of the sensitivity analyses were similar to the respective main analysis considering composite outcomes for osmotic symptoms and falls/ lower limb fractures, censoring in case of any treatment regime change, shorter follow-up time and the exclusion of patients initiating both study treatments

Results of all sensitivity analyses are given in the Appendix 5.5. Results were similar to the respective main analysis results when using composite outcomes for the osmotic symptoms and falls/ lower limb fractures or when censoring follow-up time at any change in treatment regime. Considering a shorter follow-up time of maximum 1 year did not lead to very different results than the main analysis, except that the RR results of DKA in the older patient group is not significantly increased any more. This result indicates DKA events where developed later following treatment initiation. The sensitivity analysis for which all individuals initiating both study treatments were excluded led to similar causal estimates for treatment effect and AE outcomes.

Adverse effects	n event DPP4i	n event SGLT2i	IR (CI 95%) DPP4i	IR (CI 95%) SGLT2i	RR (CI 95%)
Genital infection					
<70 years	12596	18493	102.96 (102.9, 103.02)	195.81 (195.72, 195.9)	2.27 (2.03, 2.53)
≥70 years	5406	2655	95.69 (95.61, 95.77)	195.35 (195.12, 195.59)	2.16 (1.77, 2.63)
Micturition control					
<70 years	1707	1344	12.47 (12.45, 12.49)	11.34 (11.32, 11.36)	0.64 (0.45, 0.91)
≥70 years	1924	375	31.2 (31.16, 31.25)	22.69 (22.62, 22.77)	0.81 (0.55, 1.2)
Volume depletion/ dehydration					
<70 years	1531	1246	11.17 (11.15, 11.18)	10.5 (10.49, 10.52)	0.69 (0.47, 1)
≥70 years	1391	317	22.38 (22.35, 22.42)	19.13 (19.07, 19.2)	1 (0.65, 1.56)
Urinary frequency					
<70 years	1447	1226	10.55 (10.53, 10.56)	10.34 (10.32, 10.36)	1 (0.68, 1.46)
≥70 years	1252	296	20.11 (20.08, 20.15)	17.89 (17.83, 17.96)	0.58 (0.36, 0.92)
Falls					
<70 years	2992	2253	22.08 (22.05, 22.1)	19.17 (19.15, 19.2)	0.86 (0.66, 1.13)
≥70 years	6093	1016	106.62 (106.53, 106.7)	64 (63.87, 64.12)	0.56 (0.45, 0.7)
Amputation					
<70 years	192	199	1.38 (1.38, 1.39)	1.66 (1.65, 1.67)	0.58 (0.22, 1.53)
≥70 years	117	34	1.83 (1.82, 1.84)	2.02 (1.99, 2.04)	1.14 (0.29, 4.57)
DKA					
<70 years	125	255	0.9 (0.9, 0.91)	2.13 (2.12, 2.13)	1.12 (0.41, 3.04)
≥70 years	112	50	1.75 (1.74, 1.76)	2.96 (2.93, 2.99)	3.82 (1.12, 13.03)

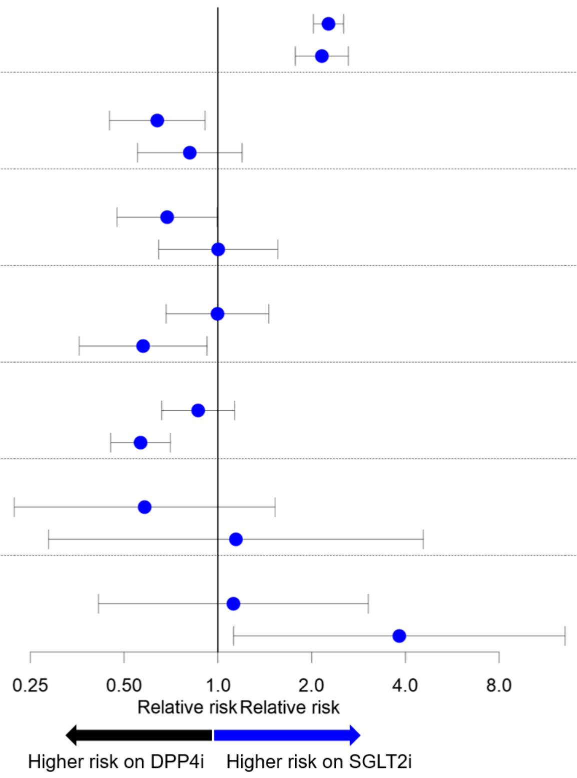


Figure 5.2: Causal effect estimation results of the relative risk (RR) for adverse effects. Additionally, the figure shows number (n) of events recorded and incidence rates (IR) per 1000 person-years. Values in brackets represent 95% confidence intervals.

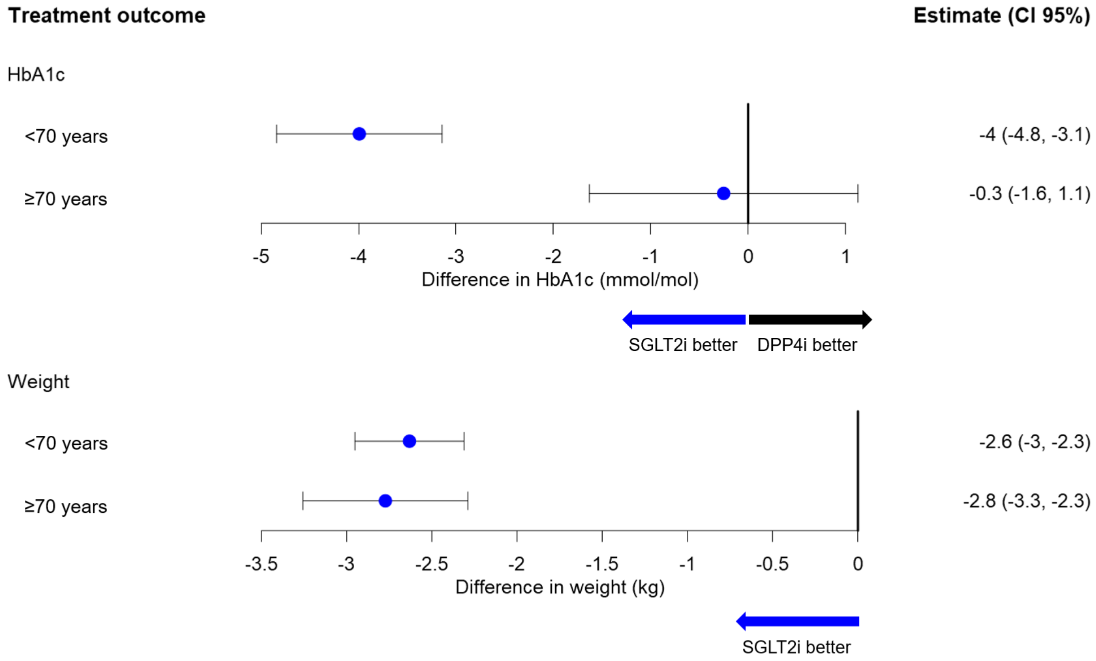


Figure 5.3: Causal effect estimation results for change in HbA1c (mmol/mol) and weight (kg). Point estimates represent the difference in outcome with SGLT2i compared to DPP4i, with negative values representing a greater HbA1c/weight reduction with SGLT2i over DPP4i.

5.5 Discussion

Our study results give confidence that SGLT2is are generally safe for older type 2 diabetes patients but AEs of concern are genital infections and DKA. Additionally, our results show that SGLT2i is effective in reducing HbA1c and weight in type 2 diabetes patients who are older than 70 years.

Relative risk for genital infections was generally increased on SGLT2i but was similar in both age groups. This confirms previous meta-analysis results [181] and results from an observational study [157], but adds to the studies with a wider age range as previous studies included too little data on older patients to draw conclusions for this patient group. DKA is the only adverse effect for which the relative risk on SGLT2i is elevated in the older population but recorded numbers were low. This finding supports the warnings of the FDA [189] and the EMA [190] and stresses the need to take DKA risk factors into account when prescribing SGLT2i to older patients. [162, 166]

Higher reduction of HbA1c for patients on SGLT2i compared to DPP4i has been shown in RCTs [248, 249] and meta-analysis [250] and in an observational study [177]. Older patients were not considered in these studies and average age of participants was low (56.7, 55.4, 56 and 55.7 years). Our analysis does not show greater reduction of HbA1c on SGLT2i compared to DPP4i for older patients and highlights the importance of analysing this group separately. This is likely due to the association between increasing age and lower eGFR, a known predictor of attenuated glycaemic response with SGLT2i. [178] Weight reduction after SGTL2i initiation is confirmed from our results for younger and older patients. Previous RCT data meta-analysis results comparing SGLT2i and DPP4i showed a weight reduction of -2.45 kg [95% CI: -2.71, -2.19] [161]. The extent of weight reduction in our study is similar to these results.

The unadjusted analysis is based on naive calculations without taking differences of baseline characteristics between the treatment groups or unmeasured confounding factors into consideration. The different results compared to our causal analysis regarding the risk of AEs illustrate the importance of a causal analysis taking into account measured and unmeasured confounding.

Furthermore, the strength of our comprehensive causal analysis lies in the application of the IV method by Ertefaie et al. [97], which addresses possible unmeasured confounding and does not rely on complete case analysis due to missingness in measured baseline characteristics. The analysis was conducted with a large real-life primary care dataset linked to hospitalization data, that captures the most important AEs for SGLT2is with primary and secondary data. This data provided the unique opportunity to adequately study older type 2 diabetes population as seen in clinical practice. Results of this study have therefore broad generalizability and can be applied to current clinical practice.

Limitations of this study are that the analysis relies on correct clinical coding of the AEs, which can be subject to inaccuracies due to miscoding or non-coding. For example, some under-representation of genital infections might be possible as antifungal medication is available over-the-counter and can be treated without having presented to primary care. Additionally, information about the severity of the AEs was not available. [157] A limitation of the IV method is that some of the data structure assumptions made are not testable with the data at hand. Additionally, as prescription preference was not measured in the data, our analysis relies on a proxy measure, which might be subject to measurement errors. Previous similar IV analysis assessing relative effectiveness and risk of type 2 diabetes treatments in the CPRD data have found that the IV assumptions are reasonable in this setting. [83, 251]

Our study provides important real-world evidence supporting careful use of SGLT2i in older type 2 diabetes patient population. Considering safety, we establish that falls, osmotic symptoms, and amputations are not increased in older patients when treated with SGLT2i. Although risk of genital infections and rare DKA is increased with SGLT2i in keeping with previous studies, reassuringly there is no evidence of an excess risk of these adverse effects in the older patient group. In terms of effectiveness, we demonstrate that SGLT2is are similar in glucose-lowering efficacy to DPP4i in this age group, with a weight benefit consistent with that observed in the wider type 2 diabetes population.

5.6 Conclusion

In conclusion, SGLT2i in the older patient population are effective and do not increase relative risk for dehydration, falls or urinary problems. Patients initiating SGLT2is have a higher relative risk to experience a genital infection compared to patients initiating DPP4i. Furthermore, DKA is a severe adverse event of concern in older patients and risk factors such as infections should be assessed before prescription of SGLT2i. This study provides a valuable causal analysis framework for the study of patient subpopulations which are generally not examined in randomized controlled trials.

5.7 Appendices

Appendix 5.1 The Instrumental Variable approach as applied in this study

Figure 5.4 represents the assumed data structure of this observational study that are pertinent to the Instrumental Variable (IV) analysis. Arrows in the graph represent assumed causal relationships between the variables. The aim of the study is to estimate the causal effect of receiving SGLT2i versus DPP4i on the outcome(s) of interest. In particular, we assume that provider prescription preference is a suitable IV and fulfils the IV assumptions, conditional on a set of measured confounders, W. The IV assumptions are: (1) The IV must be strongly associated with the exposure given W, (2) be independent of unmeasured confounders given W and (3) not have a direct effect on the outcome of interest given W. [70] Necessary conditions for the IV assumptions to hold are that (1) between-provider variation in the use of study treatment exists, (2) patient selection/assignment to a provider is unrelated to providers' preference of the study treatment, (3) a providers' use of treatment is independent of the use of alternative treatments that affect the outcome of interest. [80, 81]. The forth assumption of monotonicity is often discussed in connection with provider preference based IVs and is necessary for the identification of the point estimate of the treatment effect. This assumption requires that if a provider treats a patient with SGLT2i, all provider with a preference for SGLT2i equal or higher than the preference of this prescribing provider will also prescribe SGLT2i. The exact formulation of this assumption depends in the construction of the proxy instrument and might be invalidated because provider have to treat against there preference, for example due to insurance policies or contraindications. Therefore, estimates from provider prescription preference based IVs need to be interpret carefully.

As provider prescription preference is not directly measured in CPRD, and the set of measured confounders contains missing data, we use a proxy variable for it following the approach of Ertefaie et al. [97].

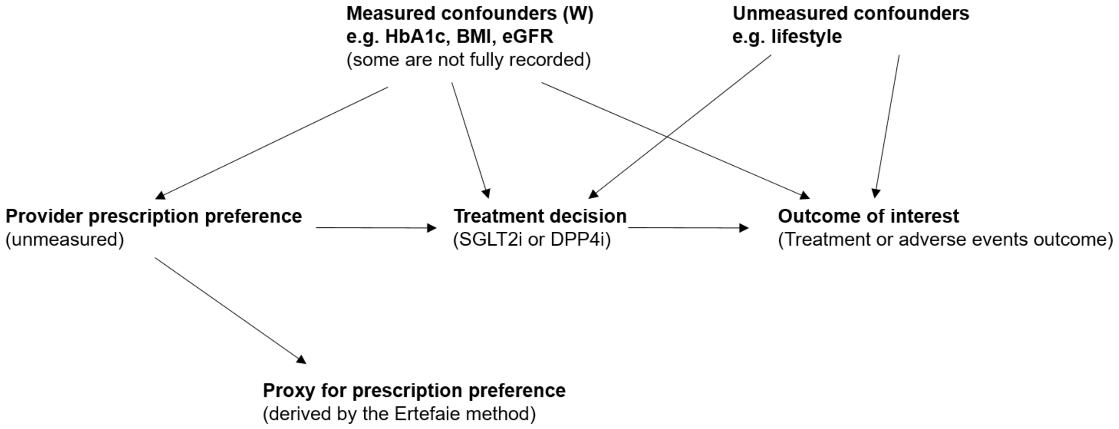

Measured confounders (W)
e.g. HbA1c, BMI, eGFR
(some are not fully recorded)
Unmeasured confounders
e.g. lifestyle
Provider prescription preference
(unmeasured)
Treatment decision
(SGLT2i or DPP4i)
Outcome of interest
(Treatment or adverse events outcome)
Proxy for prescription preference
(derived by the Ertefaie method)

Figure 5.4: Representation of assumed data structure in this study. Arrows in this plots indicate assumed relationships between variables. The missing of arrows indicates assumed lack of relationship. We assume that the variable provider prescription preference is a valid IV but it is not measured in the data at hand. Therefore, a proxy variable is derived using the IV approach by Ertefaie et al.

The following non-technical description aims to give interested readers a better understanding of the steps necessary for the construction of a proxy variable for provider prescription preference and the estimation of the causal treatment effects. For a more in depth and mathematical description of the methods, we refer to the original paper by Ertefaie et al. [97].

The Ertefaie IV approach is conducted in two steps. To apply this approach the measured confounders/ covariates are grouped into:

- W_{obs}: all measured confounders that are observed for all individuals in the data and
- W_{miss}: all measured confounders with at least one missing data point.

Step 1 of the method aims to construct a binary proxy IV for provider prescription preference which will be used as instrument, taking the value 1 when a provider has a preference for SGLT2i over DPP4i, and 0 otherwise. It is constructed using a generalized mixed effect model for the treatment decision adjusted for all measured confounders (W_{obs} and W_{miss}). The model is estimated using a complete case dataset (i.e. for individuals with complete information on W_{obs} and W_{miss}) and with a random intercept for provider. From this model the fitted values of the random intercept and its empirical distribution is used for the construction of the instrument. The instrument will take on the value 1 for each provider with an estimated random intercept larger than the median of all estimated random intercepts, and 0 otherwise. Please note that the instrument can only be calculated for provider with at least one measured covariate completely measured (W_{obs}). If this is not the case, the provider will need to be excluded from the IV analysis.

The second step of the Ertefaie method includes the calculation of the causal treatment effect with the Two-Stage Least Squares approach for continuous outcomes and the Two-Stage Predictor Substitution method otherwise. [70] This estimation step is applied to all individuals in the dataset, but with adjustment for W_{obs} only. Specifically, in stage 1, a logit model for the observed treatment decision is fitted which adjusts for W_{obs} and the instrument. Thereafter, in stage 2, the outcome is regressed on the predicted treatment decision and W_{obs} to estimate the causal treatment effect. For the continuous treatment outcomes, achieved HbA1c and weight, a linear outcome model is estimated. In case of binary adverse effect outcomes, we used a Poisson model with follow-up time (in days) as offset.

Appendix 5.2 Model summary

	Achieved HbA1c	Achieved weight	Genital infections	Osmotic symptoms	Falls	Lower limb amputations	Amputations	DKA
General patient and treatment regime information	Age, sex, ethnicity, deprivation index, smoking status, diabetes duration, year of study treatment initiation, line of therapy, number of concurrent treatments							
Biomarkers	HbA1c, eGFR, BMI/ weight, ALT							
History of comorbidities			Genital infections	Osmotic symptom, benign prostate hyperplasia	Oestrogens, oral steroids, statins, Ksparing, loops, thiazide diuretics, ACE inhibitors			
Additional treatments			Immunosuppressants, oestrogens, oral steroids	Ksparing, Loops, thiazide diuretics	lower limb fracture, falls, amputation, Diabetic ketoacidosis, Dementia, Cancer, Asthma, COPD, heart failure, CVD, CLD, osteoporosis			

Table 5.2: Summary of the relevant covariates for each outcome of interest used in the IV estimation. For the analysis of achieved weight, baseline weight instead of baseline BMI was included. Models of the first step of IV method by Ertefaie et al. and the first stage regression model of the IV estimation adjusted for all relevant covariates, whereas the outcome models only included completely recorded covariates. COPD: chronic obstructive pulmonary disease, CVD: composite of myocardial infarction, stoke, revascularisation, ischemic heart disease, angina, peripheral arterial disease, transient ischemic attack, CLD: chronic liver disease.

Appendix 5.3 Further study cohort description

	SGLT2i < 70 years (n = 66810)	SGLT2i $\geq$ 70 years (n = 10419)	DPP4i < 70 years (n = 76172)	DPP4i $\geq$ 70 years (n = 33434)
History genital infection				
Never	32233 (48.2)	5142 (49.4)	39269 (51.6)	17002 (50.9)
$<$ 1 year	8950 (13.4)	1197 (11.5)	10566 (13.9)	3920 (11.7)
1 - 5 year	13114 (19.6)	1890 (18.1)	13717 (18)	5610 (16.8)
$>$ 5 year	12513 (18.7)	2190 (21)	12620 (16.6)	6902 (20.6)
History urinary frequency				
Never	60280 (90.2)	8781 (84.3)	68673 (90.2)	28069 (84)
$<$ 1 year	819 (1.2)	178 (1.7)	1170 (1.5)	832 (2.5)
1 - 5 year	2309 (3.5)	598 (5.7)	2802 (3.7)	1969 (5.9)
$>$ 5 year	3402 (5.1)	862 (8.3)	3527 (4.6)	2564 (7.7)
History micturition control				
Never	60808 (91)	9172 (88)	69306 (91)	28375 (84.9)
$<$ 1 year	786 (1.2)	180 (1.7)	1087 (1.4)	1024 (3.1)
1 - 5 year	2024 (3)	425 (4.1)	2340 (3.1)	1732 (5.2)
$>$ 5 year	3192 (4.8)	642 (6.2)	3439 (4.5)	2303 (6.9)
History volume depletion				
Never	61180 (91.6)	9272 (89)	69803 (91.6)	28886 (86.4)
$<$ 1 year	730 (1.1)	164 (1.6)	960 (1.3)	815 (2.4)
1 - 5 year	1881 (2.8)	398 (3.8)	2173 (2.9)	1551 (4.6)
$>$ 5 year	3019 (4.5)	585 (5.6)	3236 (4.2)	2182 (6.5)
History falls				
Never	58903 (88.2)	8043 (77.2)	67251 (88.3)	24134 (72.2)
$<$ 1 year	1128 (1.7)	603 (5.8)	1473 (1.9)	2989 (8.9)
1 - 5 year	2738 (4.1)	902 (8.7)	3224 (4.2)	3634 (10.9)
$>$ 5 year	4041 (6)	871 (8.4)	4224 (5.5)	2677 (8.0)
History amputation				
Never	66477 (99.5)	10368 (99.5)	75757 (99.5)	33152 (99.2)
$<$ 1 year	75 (0.1)	6 (0.1)	111 (0.1)	62 (0.2)
1 - 5 year	147 (0.2)	20 (0.2)	182 (0.2)	115 (0.3)
$>$ 5 year	111 (0.2)	25 (0.2)	122 (0.2)	105 (0.3)

	SGLT2i	SGLT2i	DPP4i	DPP4i
	< 70 years	$\geq$ 70 years	< 70 years	$\geq$ 70 years
	(n = 66810)	(n = 10419)	(n = 76172)	(n = 33434)
History diabetic ketoacidosis				
Never	66379 (99.4)	10388 (99.7)	75805 (99.5)	33268 (99.5)
$<$ 1 year	61 (0.1)	6 (0.1)	95 (0.1)	66 (0.2)
1 - 5 year	173 (0.3)	15 (0.1)	129 (0.2)	48 (0.1)
$>$ 5 year	197 (0.3)	10 (0.1)	143 (0.2)	52 (0.2)

Table 5.3: Detailed description of potential recurrent comorbidities in the study population recorded prior to study treatment initiation.

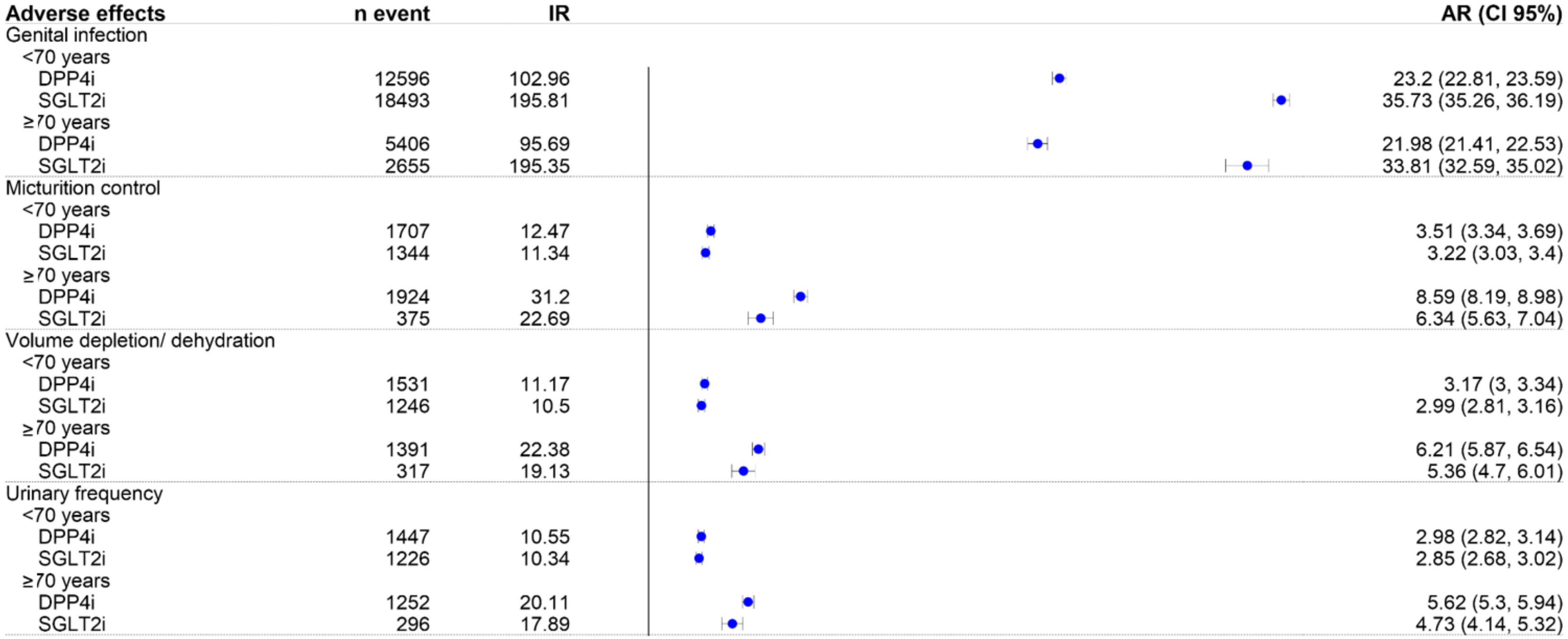
	SGLT2i	SGLT2i	DPP4i	DPP4i
	< 70 years	$\geq$ 70 years	< 70 years	$\geq$ 70 years
	(n = 66810)	(n = 10419)	(n = 76172)	(n = 33434)
HbA1c (mmol/mol)	8973 (13.4)	1406 (13.5)	7559 (9.9)	3999 (12)
eGFR ($\mathrm{ml/min/1.73m^2}$)	322 (0.5)	25 (0.2)	555 (0.7)	170 (0.5)
ALT (U/L)	4206 (6.3)	577 (5.5)	5354 (7.0)	2158 (6.5)
BMI ($\mathrm{kg/m^2}$)	2646 (4)	371 (3.6)	3701 (4.9)	1988 (5.9)
Weight (kg)	1496 (2.2)	215 (2.1)	2370 (3.1)	1440 (4.3)

Table 5.4: Summary of missing data in baseline characteristics of the study population values are given in absolute frequencies in n (%).

	Person-years of follow-up	Average follow-up time (years)
Adverse effects		
Genital infections	286867.4	1.54 (1.03)
Volume depletion	334448	1.79 (1)
Micturition control	333607.1	1.79 (1)
Urinary frequency	334596.2	1.79 (1)
Falls	326040.4	1.75 (1)
Amputation	339284.1	1.82 (0.99)
DKA	339574.1	1.82 (0.99)
Composite adverse effects		
Osmotic symptoms	329345.2	1.76 (1)
Falls + lower limb fractures	339848.6	1.82 (0.99)

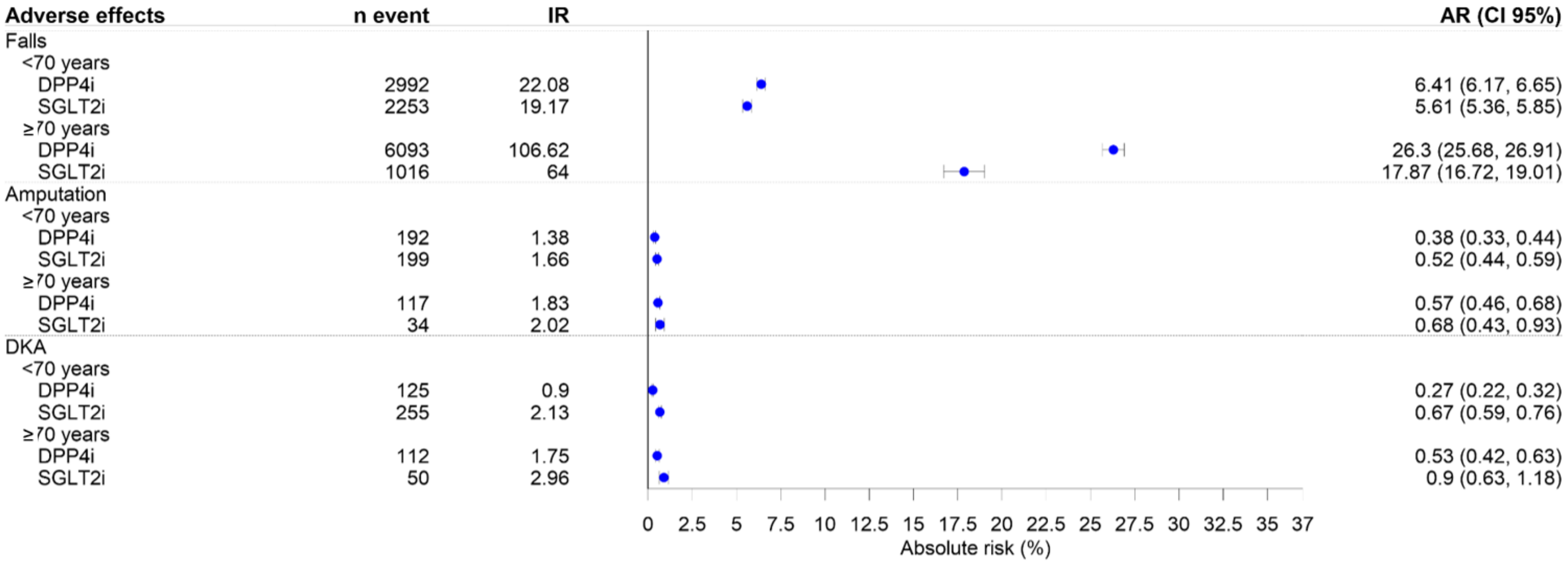
Table 5.5: Person-years of follow-up calculated as the total follow-up time and average follow-up time (in years) of the adverse effects. Values for the average follow-up time are shown as mean (standard deviation).

Appendix 5.4 Unadjusted analysis results



Adverse effects	n event	IR	AR (CI 95%)
Genital infection			
<70 years			
DPP4i	12596	102.96	23.2 (22.81, 23.59)
SGLT2i	18493	195.81	35.73 (35.26, 36.19)
≥70 years			
DPP4i	5406	95.69	21.98 (21.41, 22.53)
SGLT2i	2655	195.35	33.81 (32.59, 35.02)
Micturition control			
<70 years			
DPP4i	1707	12.47	3.51 (3.34, 3.69)
SGLT2i	1344	11.34	3.22 (3.03, 3.4)
≥70 years			
DPP4i	1924	31.2	8.59 (8.19, 8.98)
SGLT2i	375	22.69	6.34 (5.63, 7.04)
Volume depletion/ dehydration			
<70 years			
DPP4i	1531	11.17	3.17 (3, 3.34)
SGLT2i	1246	10.5	2.99 (2.81, 3.16)
≥70 years			
DPP4i	1391	22.38	6.21 (5.87, 6.54)
SGLT2i	317	19.13	5.36 (4.7, 6.01)
Urinary frequency			
<70 years			
DPP4i	1447	10.55	2.98 (2.82, 3.14)
SGLT2i	1226	10.34	2.85 (2.68, 3.02)
≥70 years			
DPP4i	1252	20.11	5.62 (5.3, 5.94)
SGLT2i	296	17.89	4.73 (4.14, 5.32)

Figure 5.5: Absolute risk (AR) results for adverse effects outcomes and the treatments SGLT2i and DPP4i, part A. The results are shown for the younger and older patient population. Additionally the figure shows number (n) of events recorded and incidence rates (IR) per 1000 person-years. Values in brackets represent 95% confidence intervals. Absolute risk results were calculated from survival model not adjusted for measured baseline characteristics.



Adverse effects	n event	IR	AR (CI 95%)
Falls			
<70 years			
DPP4i	2992	22.08	6.41 (6.17, 6.65)
SGLT2i	2253	19.17	5.61 (5.36, 5.85)
≥70 years			
DPP4i	6093	106.62	26.3 (25.68, 26.91)
SGLT2i	1016	64	17.87 (16.72, 19.01)
Amputation			
<70 years			
DPP4i	192	1.38	0.38 (0.33, 0.44)
SGLT2i	199	1.66	0.52 (0.44, 0.59)
≥70 years			
DPP4i	117	1.83	0.57 (0.46, 0.68)
SGLT2i	34	2.02	0.68 (0.43, 0.93)
DKA			
<70 years			
DPP4i	125	0.9	0.27 (0.22, 0.32)
SGLT2i	255	2.13	0.67 (0.59, 0.76)
≥70 years			
DPP4i	112	1.75	0.53 (0.42, 0.63)
SGLT2i	50	2.96	0.9 (0.63, 1.18)

Figure 5.6: Absolute risk (AR) results for adverse effects outcomes and the treatments SGLT2i and DPP4i, part B. The results are shown for the younger and older patient population. Additionally the figure shows number (n) of events recorded and incidence rates (IR) per 1000 person-years. Values in brackets represent 95% confidence intervals. Absolute risk results were calculated from survival model not adjusted for measured baseline characteristics.

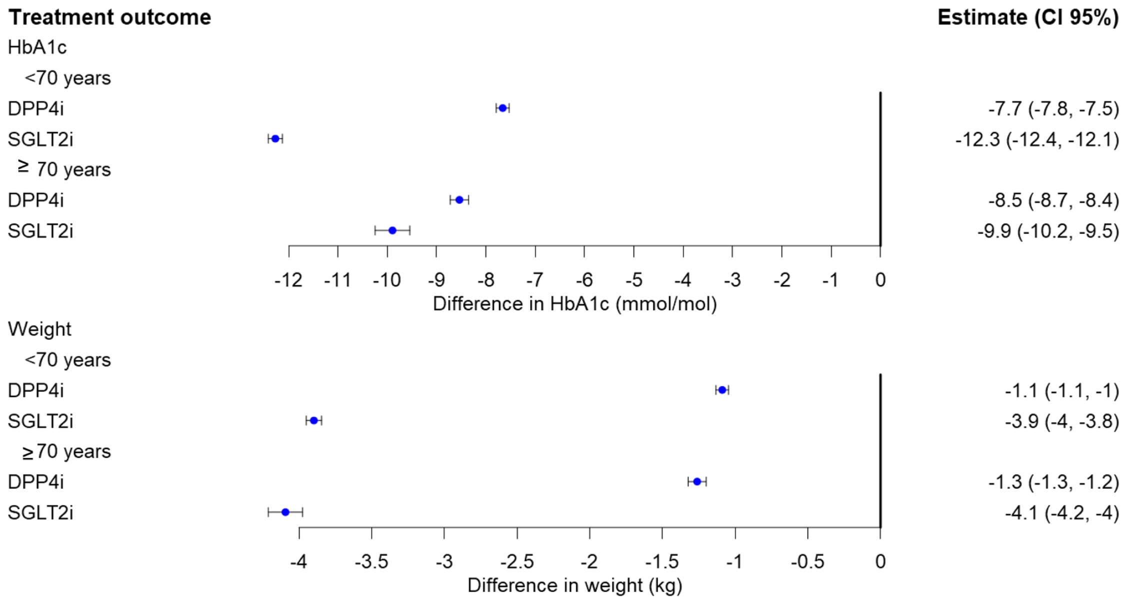


Figure 5.7: Unadjusted analysis results of the average achieved change for treatment outcomes HbA1c (mmol/mol) and weight (kg) after treatment initiation. Estimates are shown together with their 95% confidence intervals.

Appendix 5.4 Results of the sensitivity analysis

	Composite outcomes	Censoring scheme	Follow-up time	Second study period
Treatment outcomes				
Average relative difference (CI 95%)				
HbA1c (mmol/mol)				
<70 years				-4.3 (-5.1, -3.5)
≥70 years				-0.3 (-1.8, 1.2)
Weight (kg)				
<70 years				-2.7 (-2.9, -2.3)
≥70 years				-2.7 (-3.2, -2.2)
Adverse effects outcomes				
Relative risk (CI 95%)				
Genital infection				
<70 years		2.42 (2.14, 2.74)	2.55 (2.24, 2.9)	2.11 (1.9, 2.34)
≥70 years		2.16 (1.74, 2.68)	2.0 (1.59, 2.56)	2.04 (1.66, 2-51)
Micturition control				
<70 years		0.69 (0.46, 1.02)	0.72 (0.45, 1.16)	0.73 (0.52, 0.98)
≥70 years		0.91 (0.58, 1.42)	0.91 (0.54, 1.51)	1.03 (0.68, 1.54)
Volume depletion/ dehydration				
<70 years		0.73 (0.48, 1.11)	0.8 (0.49, 1.32)	0.73 (0.52, 1.03)
≥70 years		1.1 (0.67, 1.82)	1.33 (0.74, 2.38)	1.35 (0.85, 2.14)

	Composite outcomes	Censoring scheme	Follow-up time	Second study period
Urinary Frequency				
<70 years		0.99 (0.64, 1.52)	0.95 (0.57, 1.58)	1.33 (0.94, 1.88)
≥ 70 years		0.57 (0.33, 0.97)	0.75 (0.4, 1.39)	0.57 (0.34, 0.94)
Osmotic symptoms (composite)				
<70 years	0.81 (0.62, 1.06)			
≥ 70 years	0.82 (0.6, 1.12)			
Falls				
<70 years		0.83 (0.61, 1.13)	0.89 (0.61, 1.3)	0.95 (0.74, 1.2)
≥ 70 years		0.62 (0.48, 0.8)	0.46 (0.34, 0.61)	0.62 (0.49, 0.79)
Falls/ lower limb fractions (composite)				
<70 years	0.88 (0.68, 1.13)			
≥ 70 years	0.59 (0.48, 0.74)			
Amputation				
<70 years		0.91 (0.26, 3.17)	0.22 (0.05, 0.92)	0.84 (0.36, 2.0)
≥ 70 years		0.72 (0.1, 5.1)	1.4 (0.17, 11.79)	1.08 (0.26, 4.53)

	Composite outcomes	Censoring scheme	Follow-up time	Second study period
DKA				
<70 years		2.5 (0.37, 17.0)	2.46 (0.54, 11.19)	1.8 (0.72, 4.53)
≥70 years		12.05 (1.6, 91.34)	1.7 (0.3, 9.63)	6.19 (1.78, 21.5)

Table 5.6: Summary of the age-stratified sensitivity analysis results. Results for treatment outcomes are given in average relative difference in outcome measure and AE results as presented as relative risk. All results are shown with 95% confidence intervals. Composite outcomes: analysis using a composite for osmotic symptoms as well as falls and lower limb fractures. Censoring scheme: censoring for the AE is done in case of any change of the treatment regime. Follow-up time: follow-up time of the AEs was maximum 1 year. Second study period: exclusion of the second study period for patients that initiated both treatments.

Appendix 5.6 Triangulation analysis

In addition to the main analysis outlined in this chapter, the relative risks of the assorted adverse effects and comparative weight and HbA1c benefit are estimated using multivariable regression models with and without propensity score matched data (MVR and PSM respectively), as well as the Instrumental Variable method constructing the proxy instrument for prescription preference according to Brookhart et al. [92] (IV prevpatient). For this Instrumental Variable method the value of Z for each individual takes on the value of the previous treatment decision made by the same treating healthcare provider. Since values for this instrument cannot be calculated for the first treated patient within each healthcare provider, this data needs to be excluded from the analysis. Furthermore, this IV method relies on a complete case analysis with respect to data for measured confounders. As this construction method using only one previous prescription for the construction of the proxy variable, it is capable of accounting for change in provider preference. [92, 100]

The results for all adverse effects are shown in Figure 5.8 and Figure 5.9 for all methods including the main analysis results (IV ePP), which are repeated for comparison. For the osmotic symptoms, the results for MVR and PSM are consistent with the main analysis results and show increased relative risk for SGLT2is of genital infections in both age groups, but no increased relative risk for poor micturition control, volume depletion/ dehydration or urinary frequency. Results for IV prevpatient show much larger confidence intervals compared to any other estimation method and additionally conclude elevated risk of poor micturition control and volume depletion/ dehydration for the older patient population. Table 5.7 summarizes the F-statistic results for both IV methods and patient population to test the strength of both instruments. Both construction methods for a preference-based instrument lead to strong instruments with F-statistic greater than 10 for all models. A partial explanation for the larger CI of the IV prevpatient estimate is the exclusion of data for the first patient of each healthcare provider, but a more in depth analysis of the significantly elevated risk results for poor micturition control and volume depletion/ dehydration will be necessary before deriving clinical conclusions.

For the adverse effects falls, amputation and DKA the estimation results of MVR, PSM and IV prevpatient are generally consistent with the main analysis results. No elevated relative risk for all three adverse effects in neither of the patient groups can be found. Interestingly, the risk results for the older patient population and DKA are not significantly increased for MVR, PSM and IV prevpatient. Given that DKA is a rare event in the study population and the large CIs of the IV ePP results, this might be taken as an additional indication that SGLT2i is safe in the older patient population. Further analysis of the study population would be helpful in order to understand if the differing results for DKA between the estimation methods origins from the use of a complete case dataset, which all methods except IV ePP rely on.

Estimation results for HbA1c (mmol/mol) and weight (kg) are given in Figure 5.10 and are generally consistent with the main analysis results. For HbA1c and the older patient population results of MVR, PSM and IV prevpatient show a larger achieved reduction for patients on SGLT2i compared to DPP4i, but this reduction is smaller than for the patient group of younger adults.

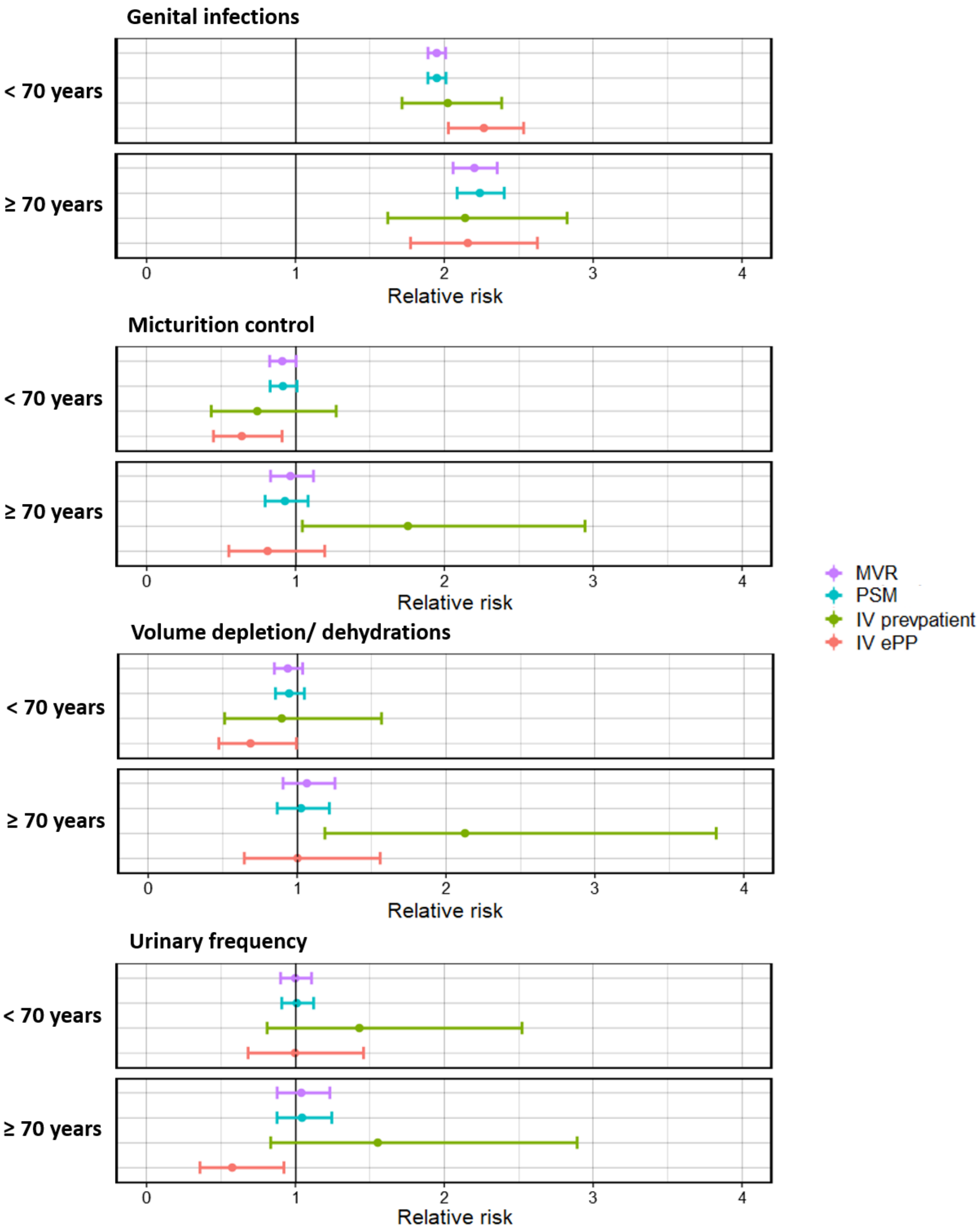


Figure 5.8: Triangulation results for osmotic symptoms adverse effects. Results are given as relative risk with 95% CIs and summarized for the methods MVR: multivariable regression, PSM: propensity score matching, IV prevpatient: IV using preference-based instrument proposed by Brookhart et al., IV ePP: IV using preference-based instrument proposed by Ertefaie et al. (main analysis)

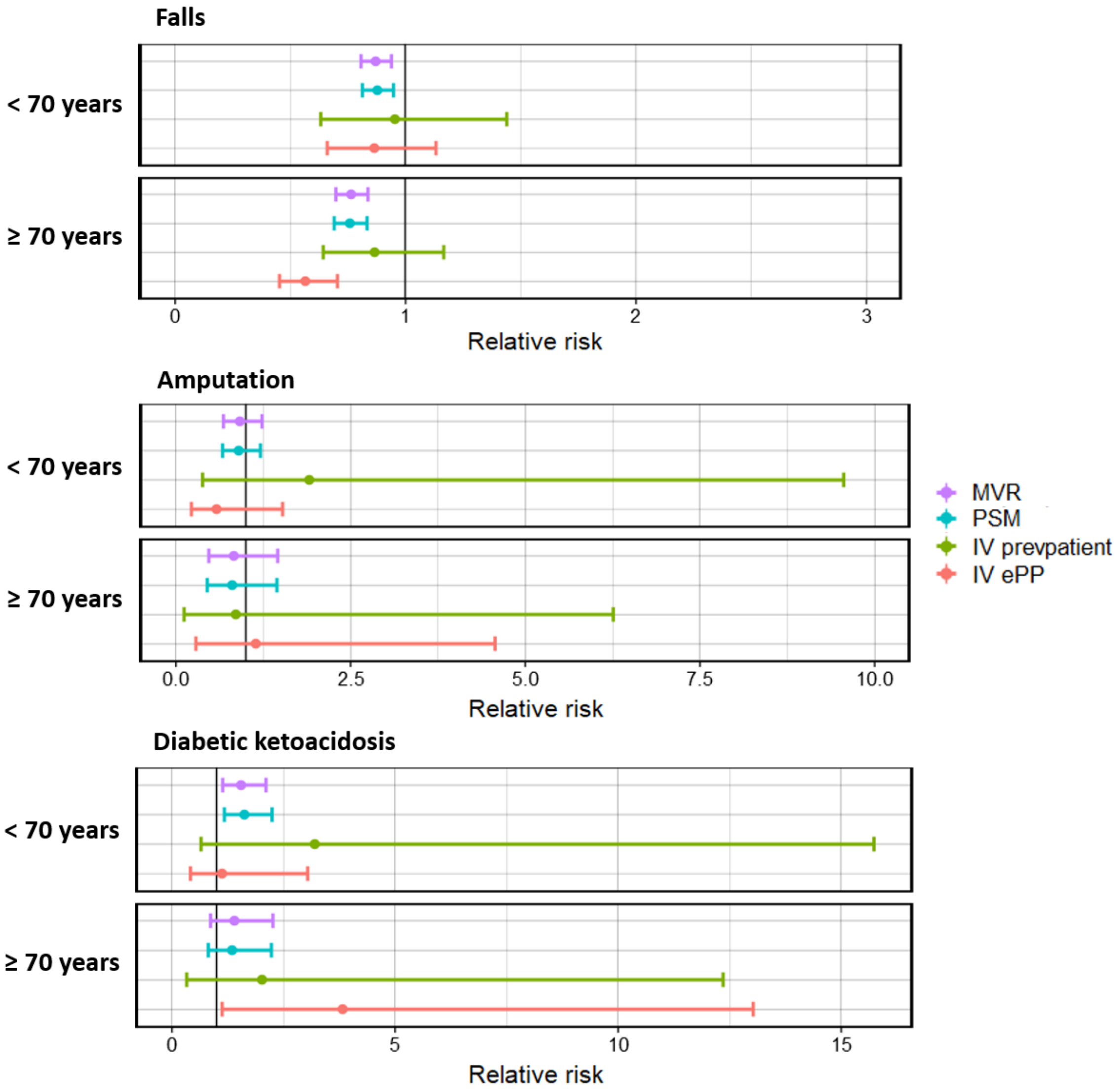


Figure 5.9: Triangulation results for falls, amputation and diabetic ketoacidosis. Results are given as relative risk with 95% CIs and summarized for the methods MVR: multivariable regression, PSM: propensity score matching, IV prevpatient: IV using preference-based instrument proposed by Brookhart et al., IV ePP: IV using preference-based instrument proposed by Ertefaie et al. (main analysis)

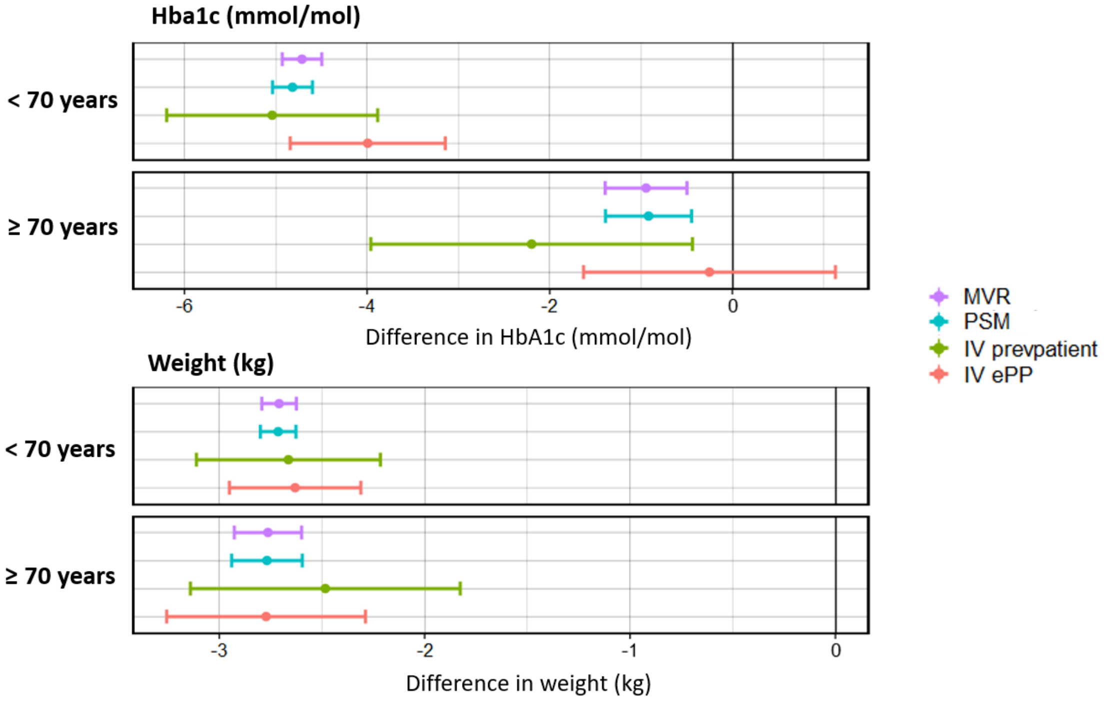


Figure 5.10: Triangulation results for HbA1c (mmol/mol) and weight (kg). Results are shown with 95% CIs and summarized for the methods MVR: multivariable regression, PSM: propensity score matching, IV prevpatient: IV using preference-based instrument proposed by Brookhart et al., IV ePP: IV using preference-based instrument proposed by Ertefaie et al. (main analysis)

		Patient population	
Outcome	IV method	Younger adults	Older adults
Genital infection	IV prevpatient	2749.76	471.22
	IV ePP	6569.15	1652.22
Micturition control	IV prevpatient	2759.81	470.79
	IV ePP	6576.54	1623.24
Volume depletion	IV prevpatient	2759.81	470.79
	IV ePP	6576.54	1623.24
Urinary frequency	IV prevpatient	2759.81	470.79
	IV ePP	6576.54	1623.24
Falls	IV prevpatient	2756.97	469.88
	IV ePP	6537.15	1631.56
Amputation	IV prevpatient	2756.97	469.88
	IV ePP	6537.15	1631.56
Diabetic ketoacidosis	IV prevpatient	2756.97	469.88
	IV ePP	6537.15	1631.56
HbA1c	IV prevpatient	2166.46	344.06
	IV ePP	5250.48	1437.66
Weight	IV prevpatient	2069.59	328.09
	IV ePP	4778.51	1284.11

Table 5.7: F-statistic results for the first stage model of the Instrumental Variable methods.

Chapter 6

Discussion

Summary

The work presented in this thesis demonstrates a causal estimation framework to derive high quality evidence from observational data valuable to guide effective treatment decisions for type 2 diabetes management. In the first two chapters the theory of causal inference, estimation methods employed in this thesis and the ideas of evidence integration from randomized controlled trials and observational studies, as well as triangulation were introduced. Limitations of randomized controlled trials and their implications to provide evidence for T2D treatment guidelines, especially for the large patient population of older adults are discussed in detail. As a main conclusion from this discussion, evidence from observational studies using for example primary care health records from the CPRD database were presented as an important addition to trial evidence on the benefit and safety profile of oral T2D treatments in populations that are often not considered in RCTs. High quality real-life evidence is therefore important to improve individualized treatment decisions for older T2D patients.

The first study presented a triangulation framework for analysing the consistency of treatment effect estimates from assorted methods which make use of different parts of the data at hand and rely on different data structure assumptions. This included the introduction of the novel prior outcome augmented Instrumental Variable approach and the generalized heterogeneity statistic to decide if two or more estimators are sufficiently similar for comparison.

The second study presented a comprehensive overview and discussion of the Instrumental Variable method using provider prescribing preference as instrument and different construction methods for a proxy instrument. A state of the art simulation study evaluated the estimation performance of the construction methods under important data conditions such as change in provider preference over time, missing data for baseline characteristics and different provider sizes. We also proposed an extended construction method which demonstrated to be capable to address change in provider preference over time and non-ignorable missingness in baseline characteristics.

On the base of this simulation study, we found a robust and established prescription preference-based Instrumental Variable method which was applied to study the relative safety and benefit profile of SGLT2i compared to DPP4i for the patient population of older adults ($\geq$ 70 years). This study evaluated important treatment outcomes and the relative risk for adverse effects which are of concern for older adults and suspected to be associated with SGLT2i such as osmotic symptoms, falls, amputations and diabetic ketoacidosis.

This chapter gives an overview of the main findings of this thesis and discusses the conclusions, implications, limitations and future research potential of each study.

Chapter 3: Triangulating Instrumental Variable, confounder adjustment and difference-in-difference methods for comparative effectiveness research in observational data

Summary

In this chapter we proposed an estimation framework for the comparative effectiveness of two treatments based on several estimation methods. This included the explanation of established confounder adjustment methods such as multivariable regression models which operate under the no unmeasured confounding assumption and causal inference methods that address unmeasured confounding such as the difference-in-difference and the Instrumental Variable method. Employing these methods in a triangulation framework for observational evidence is of value to generate high quality observational evidence, as they make use of different parts of the data and rely on different data structure assumptions. The CaT estimate relies on the adjustment of measured confounding of the study period, closely measured at time of study treatment initiation. The DiD method utilized data from the prior and study period and the Instrumental Variable method/ Control Function approach extracts variation of the treatment variable independent of confounding using a suitable instrument. We showcased these methods and their estimation performance under different data conditions in a simulation study under violations of important assumptions. The scenarios included data generation with and without unmeasured confounding, influence of the prior outcome on the study treatment decision and a direct effect of the instrument on the study outcome. Furthermore, the prior outcome augmented Instrumental Variable approach was proposed and the estimation performance of this method was shown in a simulation study. Building on work by Bowden et al. [11], we introduced the generalized Q_e statistic to formally assess if estimation methods compared in the triangulation framework target the same underlying quantity.

We showcased the triangulation framework including the POA-IV/ POA-CF method, and the Q_e statistic in an application case study on the relative risk of experiencing a genital infection on SGLT2i compared to DPP4i in a large T2D cohort. This case study was applied on data from CPRD Gold and included patients initiating SGLT2i or DPP4i as second-line treatment. As the female gender has been identified in previous observational studies as risk factors for genital infections [157], a sex-stratified analysis was conducted applying the assorted estimation procedures.

Conclusions

Simulation results showed that the POA-IV approach is able to estimate the treatment effect of interest without bias in scenarios where the IV and DiD assumptions are violated such that the prior outcome affects the treatment decision and Z influences the study outcome directly. Thereby, the POA-IV approach relies on the assumption that the new instrument from the interaction of the prior outcome and Z is unconfounded and has a strong influence on the treatment decision.

The application case study exemplified the usefulness of discussing the consistency of estimation results from different causal methods in a triangulation framework and shed light on possible sources of bias due to assumption violations of the POA-IV approach. The Q_e statistic was derived using bootstrapping and showed agreement of all estimation results except for the pairwise comparison for CaT, PSM and POA-IV, POA-CF. We hope this statistic could be a useful tool for triangulating findings from a set of distinct causal estimation strategies. Overall the estimation results showed an increased relative risk of experiencing a genital infection on SGLT2i compared to DPP4i. This results supports previous RCTs and observational evidence. [152, 154, 179] The sex-stratified analysis showed slightly elevated but not significantly different risk of females on SGLT2i for most of the estimates except the POA-IV/ POA-CF approach.

Implications

The proposed triangulation framework and Q_e statistic are helpful tools to derive high quality evidence from observational data. Discussing the consistency of estimation results in this manner provides the opportunity to analyse potential violations of data structure assumptions of the different estimation methods and hence a better understanding of the estimation results. This is important to gain more robust evidence on the benefit and safety profiles of T2D treatments which is suitable to integrate with RCT evidence for treatment guidelines. Our proposed POA-IV approach relies on a different set of data structure assumptions as the DiD and IV/ CF approach and is therefore a valuable addition to the triangulation framework.

Limitations

The proposed triangulation framework encompasses results of several estimation methods. Clinicians might find it difficult to interpret conflicting evidence and judge potential sources of bias from advanced causal inference methods such as the instrumental variable method. [91] For the application case study outlined in this chapter, coding errors of the outcome of interest might be possible in CPRD. It is also possible that not all adverse effects are recorded in the data, as less severe genital infection are treatable with over-the-counter medications. These adverse event incidences will not be recorded in primary care data. Furthermore, the study relies on a complete case analysis, which can result in bias of the treatment effect estimation in case of a informative selection of the study population. [45]

Subsequent work

To further the development of the proposed triangulation framework, we aim to develop a rigorous hierarchical testing procedure for performing a similarity analysis across an arbitrary number of estimates, whilst controlling the overall family wise error rate for the correction of multiple testing. It is also possible to extent this framework by adding other causal inference estimation methods applicable to the research question and data structure. Qualitative synthesis of estimates are being developed [8] and further approaches on ways to combine estimates will make it easier to communicate triangulation results. There is potential to combine estimates to a single more precise estimate, if they are weakly correlated but appear to be similar, as this gives credence that they are estimating the same underlying quantity. [11]

The application case study of the triangulation framework can be extended to different T2D outcomes. Long-term cardiovascular outcomes are of interest to be studied in observational data, as these outcomes are not well studied in RCTs, but important to consider for the management of T2D. Recently published observational studies, as for example Xie et al. [252], have focused on cardiovascular outcomes but do not employ causal inference methods like the Instrumental Variable approach to address unmeasured confounding. The application case study outlined in this chapter was executed using a proxy variable for healthcare provider prescription preference. The construction method for this proxy variable was proposed by Brookhart et al. [92] and used the previous prescription as value for Z. Other methods for the construction of Z are available in the literature and might be better suited for this analysis. For example, this analysis relies on a complete case dataset excluding any records with missing information on baseline characteristics. Ertefaie et al. [97] proposed an construction method for a preference-based instrument that is able to work with non-ignorable missingness in covariates without relying on a complete case analysis and could be applied for this study in order to avoid bias due to informative sample selection.

Chapter 4: Just what the doctor ordered: An evaluation of provider preference-based Instrumental Variable methods in observational studies, with application for comparative effectiveness of type 2 diabetes therapy

Summary

In this chapter we outlined a comprehensive summary of proposed rule- and model-based construction methods for a provider preference-based instrument proxy variable and compared their estimation performance in a state of the art simulation study under different data situations. The scenarios of this simulation study included different provider sizes with respect to treated patients within each provider, missing mechanism for missingness in measured confounders and structures of change in provider preference over time. Furthermore, we proposed a novel model-based construction method that utilizes a mixed effect model with random intercept of provider and a random slope for prescription time. This method aims to extend the construction method proposed by Ertefaie et al. [97] and makes it possible to address non-ignorable missingness in measured confounders as well as change in prescription preference over time for the treatment effect estimation.

All construction methods were showcased in an application case study for the estimation of the treatment effect of SGLT2i (versus DPP4i) on the reduction of achieved HbA1c (mmol/mol). We used routine data from the Clinical Practice Research Datalink (CPRD) Aurum (download November 2021) for this case study.

In this chapter we also revised the application case study outlined in Chapter 3, summarized in Appendix 4.8 and applied the Ertefaie method for constructing a preference-based proxy instrument for the relative risk analysis of experiencing a genital infection on SGLT2i.

Conclusions

The state-of-the art simulation study showed that the more complex model-based construction methods require enough data within each provider for good estimation performance and can lead to biased results in case of small provider sample sizes. Rule-based methods were able to better accommodate smaller provider sample sizes. But not all rule-based methods performed well in case of changing provider preference. The construction methods which rely all previous patient prescriptions struggled to reflect on the change in preference and their estimation results were biased. Only the model-based construction method by Ertefaie et al. [97] and our extension method were able to estimate the treatment effect without bias in case of non-ignorable missingness. Both methods lead to very similar results, with our proposed method being slightly more efficient in case of change in provider preference.

In the application case study, most construction methods lead to similar results and the analysis confirmed a blood glucose lowering benefit of SGLT2is which is on average higher compared to DPP4i. The Ertefaie method and our extension method concluded a lower comparative HbA1c reduction benefit compared to all other IV methods which did not origin from a selection of patients due to complete case analysis.

The revision of the application case study outline in Chapter 3 showed similar but more efficient estimation results when employing the Ertefaie method compared to the construction method based on one previous prescription.

Implications

The choice of method to construct a proxy instrument for provider prescription preference should be made under consideration of the data structure including aspects such as the possibility of change in preference over time or missing data in baseline characteristics important for the outcome model. Model-based methods such as the Ertefaie method and our extension method are robust and suitable for the causal effect estimation if enough data within each provider is available. It can be helpful to triangulate estimation results for different model- and rule-based methods to check possible inconsistencies that could direct towards weaknesses of certain construction methods.

Provider preference-based Instrumental Variable analysis have proven to be a useful tool for investigating the benefit and risk profile of oral T2D treatments. The established Ertefaie method showed robustly good estimation performance under different data conditions and should be considered for further causal inference studies in T2D research.

Limitations

The simulation study outlined in this chapter is limited in the number of presented scenarios which provides possibilities for further research. The Ertefaie method and our proposed extension method have only been tested on one non-ignorable missingness mechanism, as we employed the same data generation strategy introduced in the original paper. Furthermore, provider sizes was chosen to be equal for all simulated providers which does not reflect well on real world GP sizes which can differ greatly, for example by regions. Lastly, the simulation showed interesting results regarding the rule-based construction methods based on previous provider prescriptions and their instrument strength which could not be fully elucidated in this study.

Subsequent work

An example for future studies are the consideration of different data generation models for non-ignorable missingness to further test the robustness of the Ertefaie method and our extension method. It will also be interesting to explore the performance of model-based construction methods in case some smaller provider are included in the data as this reflects better on the real world situation of GPs. Additional investigations regarding the rule-based methods which are based on a subset of previous prescription data are needed to further understand some of their simulation results. In the simulation outlined in this chapter these methods showed weak instrument strengths under certain data conditions, but their estimation results did not show weak instrument bias.

Chapter 5: Evaluation of the safety and effectiveness of SGLT2 inhibitors in adults over 70 using an instrumental variable approach: UK population based study

Summary

In this chapter we conducted a causal inference analysis for the comparative safety and benefit profile of SGLT2i versus DPP4i in the patient population of older adults 70 years or older. Outcomes of interest for this analysis included adverse effects which are of concern in the older patient population such as genital infections, poor micturition control, volume depletion/ dehydration, urinary frequency, falls, lower limb amputations and diabetic ketoacidosis. Treatment outcomes of interest are achieved HbA1c (mmol/mol) and weight (kg) 12 months after treatment initiation. The causal analysis was conducted utilizing the IV approach by Ertefaie et al. [97] with a proxy variable on healthcare provider prescription preference as instrument to create a pseudo-randomized sample and estimate the treatment effect under consideration of measured and unmeasured confounding. The T2D study population was selected from the CPRD Aurum database which is a large UK primary care dataset. Furthermore, CPRD data was linked to Hospital Episode statistics and data from the Office for National Statistics on death registrations, as well as patient level Index of Deprivation data. This made it possible to conduct an in depth study on a multitude of adverse effects potentially related to SGLT2i with an appropriate sample size of the subpopulation of older patients.

Results of the causal inference study provide high quality observational evidence which contribute towards closing the gap of spare evidence-based treatment guidelines for older T2D patients, due to a lack of RCT evidence from this patient population.

Conclusions

This causal analysis leverages the strengths of the Ertefaie method of constructing a preference-based instrument taking the prescription behaviour and all measured confounders into consideration to mitigate the risk of measurement errors. Additionally, the method does not rely on a complete case dataset and has the potential to estimate the causal treatment effect without bias, even in case of non-ignorable missingness. The observational evidence is based on the rich UK primary care data of CPRD which made subgroup analysis on the important and large patient population of older adults with T2D possible. The study results give confidence for the safety of treating older T2D patients with SGLT2is and confirms the glycaemic and weight loss benefits of this drug class previously found for more general patient populations in RCTs. Adverse effects of concern are genital infections and DKA. Relative risk for genital infection is increased for patients on SGLT2is, with similar elevated treatment effect estimates for younger and older adults. Relative DKA risk is increased for older T2D patients when estimated using the IV method proposed by Ertefaie et al. [97], but the triangulation results using IV prevpatient and conventional multivariable regression with and without propensity score matching did not replicate these results. Furthermore, DKA is a rare event in this study population and estimation results are based on only a few cases.

Implications

SGLT2i in the older adults are effective and generally safe regarding adverse effect of micturition control, urinary frequency, falls and lower limb amputations. When prescribing SGLT2is to older adults clinicians should consider and evaluate existing risk factors of genital infections and DKA such as previous genital infections and DKA events.

Limitations

Non coding of adverse effects in primary care data, for example genital infections treated with over-the-counter treatments is possible which will lead to an underrepresentation of weaker adverse effects. Furthermore, the Instrumental Variable method depends on partially untestable data structure assumptions regarding unmeasured confounders that need to be justified using subject matter knowledge. As provider prescription preference is not measured in the CPRD data, the Instrumental variable method applied for the main analysis and the triangulation analysis rely on utilizing the data at hand to construct a proxy variable for this preference-based instrument. Therefore, measurement errors for the proxy variable of prescription preference are possible.

Subsequent work

We developed a robust causal estimation framework for the analysis of patient subpopulations and with an Instrumental Variable method that is able to estimate treatment effects in case of unmeasured confounding and non-ignorable missingness in baseline characteristics data. This estimation framework can be applied to different treatment and adverse effect outcomes of interest for T2D research, for example long-term outcomes such as cardiovascular endpoints. Furthermore, the framework can be applied to study treatment outcomes in different ethnicity patient subpopulation. As previous studies have shown differential ethnic predisposition and pathophysiology of type 2 diabetes [253], observational evidence for different ethnicity subpopulations are important to improve individualized T2D treatment.

Bibliography

[1] M. L. Meldrum, “A brief history of the randomized controlled trial: From oranges and lemons to the gold standard,” *Hematology/Oncology Clinics of North America*, vol. 14, pp. 745–760, 2000.

[2] S. Houle, “An Introduction to the Fundamentals of Randomized Controlled Trials in Pharmacy Research,” *The Canadian Journal of Hospital Pharmacy*, vol. 68, p. 28, 2015.

[3] L. M. Friedman, C. D. Furberg, D. L. DeMets, D. M. Reboussin, and C. B. Granger, *Fundamentals of Clinical Trials*. Springer, 2015.

[4] C. Winship and S. L. Morgan, “THE ESTIMATION OF CAUSAL EFFECTS FROM OBSERVATIONAL DATA,” *Annual Review of Sociology*, vol. 25, pp. 659–706, 1999. [Online]. Available: www.annualreviews.org

[5] E. J. Boyko, “Observational research - opportunities and limitations,” *Journal of Diabetes and Its Complications*, vol. 27, pp. 642–648, 2013.

[6] S. Greenland and H. Morgenstern, “Confounding in Health Research,” *Annual Review of Public Health*, vol. 22, pp. 189–212, 2001.

[7] J. J. Heckman, “The Scientific Model of Causality,” *Sociological Methodology*, vol. 35, pp. 1–97, 2005.

[8] G. Hammerton and M. R. Munafò, “Causal inference with observational data: the need for triangulation of evidence,” *Psychological Medicine*, vol. 51, no. 4, pp. 563–578, 2021.

[9] D. A. Lawlor, K. Tilling, and G. Davey Smith, “Triangulation in aetiological epidemiology,” *International Journal of Epidemiology*, vol. 45, no. 6, pp. 1866–1886, 2016.

[10] M. R. Munafò and G. D. Smith, “Repeating experiments is not enough,” *Nature*, vol. 553, no. 7689, pp. 399–401, 2018.

[11] J. Bowden, L. C. Pilling, D. Türkmen, C.-L. Kuo, and D. Melzer, “The Triangulation WIthin a STudy (TWIST) framework for causal inference within pharmacogenetic research,” *PLoS Genetics*, vol. 17, p. e1009783, 2021.

[12] M. Höfler, “Causal inference based on counterfactuals,” *BMC Medical Research Methodology*, vol. 5, pp. 1–12, 2005.

[13] Y. Raita, C. A. C. Jr, L. Liang, and K. Hasegawa, “Big Data, Data Science, and Causal Inference: A Primer for Clinicians,” *Frontiers in Medicine*, vol. 8, p. 678047, 2021.

[14] J. Pearl, “Causal inference in statistics: An overview,” *Statistics Surveys*, vol. 3, pp. 96–146, 2009.

[15] H. Oppewal, “Concept of Causality and Conditions for Causality,” *Wiley International Encyclopedia of Marketing*, 2010.

[16] D. Hume, “An enquiry concerning human understanding (Introduction, notes, and editorial arrangement by Antony Flew),” *La Salle, IL: Open Court. (Original work published 1748)*, 1988.

[17] J. Pearl, *Causality*. Cambridge university press, 2009.

[18] K. J. Rothman, “Causal Inference,” *Epidemiology Resources Inc.*, 1988.

[19] K. J. Rothman and S. Greenland, “Causation and Causal Inference in Epidemiology,” *American Journal of Public Health*, vol. 95, pp. S144–S150, 2005.

[20] E. Igelström, P. Craig, J. Lewsey, J. Lynch, A. Pearce, and S. V. Katikireddi, “Causal inference and effect estimation using observational data,” *Journal of Epidemiology and Community Health*, vol. 76, pp. 960–966, 2022.

[21] M. A. Hernán and J. M. Robins, “Estimating causal effects from epidemiological data,” *Journal of Epidemiology & Community Health*, vol. 60, pp. 578–586, 2006.

[22] M. L. Petersen, K. E. Porter, S. Gruber, Y. Wang, and M. J. V. D. Laan, “Diagnosing and responding to violations in the positivity assumption,” *Statistical Methods in Medical Research*, vol. 21, pp. 31–54, 2012.

[23] T. J. VanderWeele, “Brief Report: Concerning the Consistency Assumption in Causal Inference,” *Epidemiology*, vol. 20, pp. 880–883, 2009.

[24] M. G. Hudgens and M. E. Halloran, “Toward Causal Inference With Interference,” *Journal of the American Statistical Association*, vol. 103, pp. 832–842, 2008.

[25] M. A. Hernán, “Beyond exchangeability: The other conditions for causal inference in medical research,” *Statistical Methods in Medical Research*, vol. 21, pp. 3–5, 2012.

[26] J. C. Digitale, J. N. Martin, and M. M. Glymour, “Tutorial on directed acyclic graphs,” *Journal of Clinical Epidemiology*, vol. 142, pp. 264–267, 2022.

[27] N. Pearce and D. A. Lawlor, “Causal inference—so much more than statistics,” *International Journal of Epidemiology*, vol. 45, pp. 1895–1903, 2016.

[28] M. Glymour, J. Pearl, and N. P. Jewell, *Causal Inference in Statistics: A Primer*. John Wiley & Sons, 2016.

[29] E. L. Hannan, “Randomized Clinical Trials and Observational Studies: Guidelines for Assessing Respective Strengths and Limitations,” *JACC: Cardiovascular Interventions*, vol. 1, pp. 211–217, 2008.

[30] C. E. McCoy, “Understanding the Intention-to-treat Principle in Randomized Controlled Trials,” *Western Journal of Emergency Medicine*, vol. 18, p. 1075, 2017.

[31] C. Y. Lu, “Observational studies: a review of study designs, challenges and strategies to reduce confounding,” *International Journal of Clinical Practice*, vol. 63, pp. 691–697, 2009.

[32] M. S. Thiese, “Observational and interventional study design types; an overview,” *Biochemia Medica*, vol. 24, pp. 199–210, 2014.

[33] D. L. Sackett, W. M. C. Rosenberg, J. A. M. Gray, R. B. Haynes, and W. S. Richardson, “Evidence based medicine: what it is and what it isn’t,” *BMJ*, vol. 312, pp. 71–72, 1996.

[34] A. L. Rosner, “Evidence-based medicine: Revisiting the pyramid of priorities,” *Journal of Bodywork and Movement Therapies*, vol. 16, pp. 42–49, 2012.

[35] T. Shaneyfelt, “Pyramids are guides not rules: the evolution of the evidence pyramid,” pp. 121–122, 2016.

[36] P. S. Mulimani, “Evidence-based practice and the evidence pyramid: A 21st century orthodontic odyssey,” *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 152, pp. 1–8, 2017.

[37] L. R. Johnston Jr, “Moving forward by looking back: “retrospective” clinical studies,” *Journal of Orthodontics*, vol. 29, pp. 221–226, 2002.

[38] W. R. Proffit, “Evidence and clinical decisions: Asking the right questions to obtain clinically useful answers,” *Seminars in Orthodontics*, vol. 19, pp. 130–136, 2013.

[39] N. K. Denzin, *The Research Act: A Theoretical Introduction to Sociological Methods*, 2nd ed. McGraw-Hill, 1978.

[40] A. B. Hill, “The Environment and Disease: Association or Causation?” 1965.

[41] P. R. Rosenbaum and D. B. Rubin, “The central role of the propensity score in observational studies for causal effects,” *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983.

[42] P. C. Austin, “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies,” *Multivariate Behavioral Research*, vol. 46, no. 3, pp. 399–424, 2011.

[43] E. W. Steyerberg, *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer, 2009.

[44] O. Baser, “Choosing propensity score matching over regression adjustment for causal inference: when, why and how it makes sense,” *Journal of Medical Economics*, vol. 10, no. 4, pp. 379–391, 2007.

[45] R. J. A. Little and D. B. Rubin, *Statistical Analysis With Missing Data*. John Wiley & Sons, Inc., 2002.

[46] K. L. Sainani, “Propensity Scores: Uses and Limitations,” *American Academy of Physical Medicine and Rehabilitation*, vol. 4, no. 9, pp. 693–697, 2012.

[47] P. C. Austin and M. M. Mamdani, “A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use,” *Statistics in Medicine*, vol. 25, no. 12, pp. 2084–2106, 2006.

[48] R. L. Tannen, M. G. Weiner, and D. Xie, “Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the ‘prior event rate ratio’ to adjust for unmeasured confounding by indication,” *Pharmacoepidemiology and Drug Safety*, vol. 17, pp. 671–685, 2008.

[49] N. X. Lin and W. E. Henley, “Prior event rate ratio adjustment for hidden confounding in observational studies of treatment effectiveness: a pairwise Cox likelihood approach,” *Statistics in Medicine*, vol. 35, pp. 5149–5169, 12 2016.

[50] L. R. Rodgers, J. M. Dennis, B. M. Shields, L. Mounce, I. Fisher, A. T. Hattersley, and W. E. Henley, “Prior event rate ratio adjustment produced estimates consistent with randomized trial: a diabetes case study,” *Journal of Clinical Epidemiology*, vol. 122, pp. 78–86, 6 2020.

[51] I. Petersen, I. Douglas, and H. Whitaker, “Self controlled case series methods: an alternative to standard epidemiological study designs,” *bmj*, vol. 354, 2016.

[52] M. Yu, D. Xie, X. Wang, M. G. Weiner, and R. L. Tannen, “Prior event rate ratio adjustment: numerical studies of a statistical method to address unrecognized confounding in observational studies,” *Pharmacoepidemiology and Drug Safety*, vol. 21, pp. 60–68, 2012.

[53] E. W. Thommes, S. M. Mahmud, Y. Young-Xu, J. T. Snider, R. van Aalst, J. K. Lee, Y. Halchenko, E. Russo, and A. Chit, “Assessing the prior event rate ratio method via probabilistic bias analysis on a Bayesian network,” *Statistics in Medicine*, vol. 39, no. 5, pp. 639–659, 2020.

[54] M. J. Uddin, R. H. Groenwold, T. P. Van Staa, A. De Boer, S. V. Belitser, A. W. Hoes, K. C. Roes, and O. H. Klungel, “Performance of prior event rate ratio adjustment method in pharmacoepidemiology: a simulation study,” *Pharmacoepidemiology and Drug Safety*, vol. 24, no. 5, pp. 468–477, 2015.

[55] A. J. Streeter, N. X. Lin, L. Crathorne, M. Haasova, C. Hyde, D. Melzer, and E. H. William, “Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review,” *Journal of Clinical Epidemiology*, vol. 87, pp. 23–34, 2017.

[56] J. B. Dimick and A. M. Ryan, “Methods for Evaluating Changes in Health Care PolicyThe Difference-in-Differences Approach,” *JAMA*, vol. 312, no. 22, pp. 2401–2402, 2014.

[57] H. Zhou, C. Taber, S. Arcona, and Y. Li, “Difference-in-Differences Method in Comparative Effectiveness Research: Utility with Unbalanced Groups,” *Applied Health Economics and Health Policy*, vol. 14, pp. 419–429, 2016.

[58] M. A. Hernán and J. M. Robins, “Instruments for Causal Inference: An Epidemiologist’s Dream?” *Epidemiology*, pp. 360–372, 2006.

[59] T. Widding-Havneraas, A. Chaulagain, I. Lyhmann, H. D. Zachrisson, F. Elwert, S. Markussen, D. McDaid, and A. Mykletun, “Preference-based instrumental variables in health research rely on important and underreported assumptions: a systematic review,” *Journal of Clinical Epidemiology*, vol. 139, pp. 269–278, 2021.

[60] J. D. Angrist, G. W. Imbens, and D. B. Rubin, “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 444–455, 1996.

[61] P. G. Wright, *The tariff on animal and vegetable oils*. Macmillan, 1928.

[62] A. Balke and J. Pearl, “Bounds on Treatment Effects from Studies with Imperfect Compliance,” *Journal of the American Statistical Association*, vol. 92, no. 439, pp. 1171–1176, 1997.

[63] S. A. Swanson and M. A. Hernán, “Commentary: How to Report Instrumental Variable Analyses (Suggestions Welcome),” *Epidemiology*, vol. 24, no. 3, pp. 370–374, 2013.

[64] J. Labrecque and S. A. Swanson, “Understanding the Assumptions Underlying Instrumental Variable Analyses: a Brief Review of Falsification Strategies and Related Tools,” *Current Epidemiology Reports*, vol. 5, pp. 214–220, 2018.

[65] D. S. Small, Z. Tan, R. R. Ramsahai, S. A. Lorch, and M. A. Brookhart, “Instrumental Variable Estimation with a Stochastic Monotonicity Assumption,” *Statistical Science*, vol. 32, pp. 561–579, 2007.

[66] S. A. Swanson and M. A. Hernán, “Think globally, act globally: an epidemiologist’s perspective on instrumental variable estimation,” *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, vol. 29, no. 3, p. 371, 2014.

[67] S. A. Swanson, M. Miller, J. M. Robins, and M. A. Hernán, “Definition and Evaluation of the Monotonicity Condition for Preference-Based Instruments,” *Epidemiology*, vol. 26, p. 414, 2015.

[68] S. Aso and H. Yasunaga, “Introduction to Instrumental Variable Analysis,” *Annals of Clinical Epidemiology*, vol. 2, pp. 69–74, 2020.

[69] O. Klungel, M. J. Uddin, A. de Boer, S. Belitser, R. Groenwold, K. Roes *et al.*, “Instrumental Variable Analysis in Epidemiologic Studies: An Overview of the Estimation Methods,” *Pharmaceutica Analytica Acta*, vol. 6, no. 353, p. 2, 2015.

[70] M. L. Lousdal, “An introduction to instrumental variable assumptions, validation and estimation,” *Emerging Themes in Epidemiology*, vol. 15, pp. 1–7, 1 2018.

[71] J. M. Wooldridge, *Introductory Econometrics - A Modern Approach*, 7th ed. South-Western College Publishing, 2019.

[72] R. J. Bowden and D. A. Turkington, *Instrumental Variables*. Cambridge: Cambridge University Press, 1984.

[73] J. Garen, "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable," *Econometrica: Journal of the Econometric Society*, pp. 1199–1218, 1984.

[74] J. M. Wooldridge, "On two stage least squares estimation of the average treatment effect in a random coefficient model," *Economics Letters*, vol. 56, no. 2, pp. 129–133, 1997.

[75] E. T. Tchetgen, "A Note on the Control Function Approach with an Instrumental Variable and a Binary Outcome," *Epidemiologic Methods*, vol. 3, pp. 107–112, 2014.

[76] E. A. Vertosick, M. Assel, and A. J. Vickers, "A systematic review of instrumental variable analyses using geographic region as an instrument," *Cancer Epidemiology*, vol. 51, pp. 49–55, 2017.

[77] J. Stock, J. H. Wright, and M. Yogo, "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, vol. 20, pp. 518–529, 2002.

[78] M. M. Glymour, E. J. Tchetgen Tchetgen, and J. M. Robins, "Credible Mendelian Randomization Studies: Approaches for Evaluating the Instrumental Variable Assumptions," *American Journal of Epidemiology*, vol. 175, no. 4, pp. 332–339, 2012.

[79] J. A. Rassen, M. A. Brookhart, R. J. Glynn, M. A. Mittleman, and S. Schneeweiss, "Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships," *Journal of Clinical Epidemiology*, vol. 62, no. 12, pp. 1226–1232, 2009.

[80] E. L. Korn and S. Baumrind, "Clinician Preferences and the Estimation of Causal Treatment Differences," *Statistical Science*, vol. 13, pp. 209–235, 1998.

[81] M. A. Brookhart and S. Schneeweiss, “Preference-Based Instrumental Variable Methods for the Estimation of Treatment Effects: Assessing Validity and Interpreting Results,” *The International Journal of Biostatistics*, vol. 3, 2007.

[82] R. Ionescu-Ittu, J. A. C. Delaney, and M. Abrahamowicz, “Bias–variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study,” *Pharmacoepidemiology and Drug Safety*, vol. 18, pp. 562–571, 2009.

[83] P. Bidulka, S. O’Neill, A. Basu, S. Wilkinson, R. J. Silverwood, P. Charlton, A. Briggs, A. I. Adler, K. Khunti, L. A. Tomlinson, L. Smeeth, I. J. Douglas, and R. Grieve, “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*, vol. 11, 9 2021.

[84] C. A. Emdin, A. J. Hsiao, A. Kiran, N. Conrad, G. Salimi-Khorshidi, M. Woodward, S. G. Anderson, H. Mohseni, J. J. V. McMurray, and J. G. F. Cleland, “Referral for Specialist Follow-up and Its Association With Post-discharge Mortality Among Patients With Systolic Heart Failure (from the National Heart Failure Audit for England and Wales),” *The American Journal of Cardiology*, vol. 119, pp. 440–444, 2017.

[85] N. Pratt, E. E. Roughead, P. Ryan, and A. Salter, “Antipsychotics and the risk of death in the elderly: an instrumental variable analysis using two preference based instruments,” *Pharmacoepidemiology and Drug Safety*, vol. 19, pp. 699–707, 2010.

[86] S. Dalsgaard, H. S. Nielsen, and M. Simonsen, “Consequences of ADHD medication use for children’s outcomes,” *Journal of Health Economics*, vol. 37, pp. 137–151, 2014.

[87] A. G. Boef, J. van Paassen, M. S. Arbous, A. Middelkoop, J. P. Vandenbroucke, S. le Cessie, and O. M. Dekkers, “Brief Report: Physician’s Preference-based Instrumental Variable Analysis: Is It Valid and Useful in a Moderate-sized Study?” *Epidemiology*, pp. 923–927, 2014.

[88] N. M. Davies, D. Gunnell, K. H. Thomas, C. Metcalfe, F. Windmeijer, and R. M. Martin, “Physicians’ prescribing preferences were a potential instrument for patients’ actual prescriptions of antidepressants,” *Journal of Clinical Epidemiology*, vol. 66, pp. 1386–1396, 2013.

[89] N. M. Davies, G. D. Smith, F. Windmeijer, and M. R. M., “Issues in the Reporting and Conduct of Instrumental Variable Studies: A Systematic Review,” *Epidemiology*, vol. 24, pp. 363–369, 2013.

[90] A. G. C. Boef, S. le Cessie, O. M. Dekkers, P. Frey, P. M. Kearney, N. Kerse, C. D. Mallen, V. J. C. McCarthy, S. P. Mooijaart, and C. Muth, “Physician’s Prescribing Preference as an Instrumental Variable,” *Epidemiology*, vol. 27, pp. 276–283, 2016.

[91] N. M. Davies, K. H. Thomas, A. E. Taylor, G. M. J. Taylor, R. M. Martin, M. R. Munafò, and F. Windmeijer, “How to compare instrumental variable and conventional regression analyses using negative controls and bias plots,” *International Journal of Epidemiology*, vol. 46, pp. 2067–2077, 2017.

[92] M. A. Brookhart, P. Wang, D. H. Solomon, and S. Schneeweiss, “Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable,” *Epidemiology*, vol. 17, pp. 268–275, 2006.

[93] M. Abrahamowicz, M.-E. Beauchamp, R. Ionescu-Ittu, J. A. C. Delaney, and L. Pilote, “Reducing the Variance of the Prescribing Preference-based Instrumental Variable Estimates of the Treatment Effect,” *American Journal of Epidemiology*, vol. 174, pp. 494–502, 2011.

[94] M. A. Brookhart, J. A. Rassen, and S. Schneeweiss, “Instrumental variable methods in comparative safety and effectiveness research,” *Pharmacoepidemiology and Drug Safety*, vol. 19, pp. 537–554, 2010.

[95] R. S. Stafford, C. D. Furberg, S. N. Finkelstein, I. M. Cockburn, T. Alehegn, and J. Ma, “Impact of clinical trial results on national trends in α-blocker prescribing, 1996-2002,” *Jama*, vol. 291, pp. 54–62, 2004.

[96] C. A. Jackevicius, G. M. Anderson, L. Leiter, and J. V. Tu, “Use of the Statins in Patients After Acute Myocardial Infarction Does Evidence Change Practice?” *Archives of Internal Medicine*, vol. 161, pp. 183–188, 2001.

[97] A. Ertefaie, J. H. Flory, S. Hennessy, and D. S. Small, “Instrumental Variable Methods for Continuous Outcomes That Accommodate Nonignorable Missing Baseline Values,” *American Journal of Epidemiology*, vol. 185, pp. 1233–1239, 2017.

[98] S. Hennessy, C. E. Leonard, C. M. Palumbo, X. Shi, and T. R. T. Have, “Instantaneous preference was a stronger instrumental variable than 3-and 6-month prescribing preference for NSAIDs,” *Journal of Clinical Epidemiology*, vol. 61, pp. 1285–1288, 2008.

[99] T. B. Newman, E. Vittinghoff, and C. E. McCulloch, “Efficacy of Phototherapy for Newborns with Hyperbilirubinemia: A Cautionary Example of an Instrumental Variable Analysis,” *Medical Decision Making*, vol. 32, pp. 83–92, 2012.

[100] J. A. Rassen, M. A. Brookhart, R. J. Glynn, M. A. Mittleman, and S. Schneeweiss, “Instrumental variables II: instrumental variable application—in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance,” *Journal of Clinical Epidemiology*, vol. 62, pp. 1233–1241, 2009.

[101] B. J. Potter, C. Dormuth, and J. L. Lorier, “A theoretical exploration of therapeutic monomania as a physician-based instrumental variable,” *Pharmacoepidemiology and Drug Safety*, vol. 29, pp. 45–52, 2020.

[102] R. Vieira, S. B. Souto, E. Sánchez-López, A. López Machado, P. Severino, S. Jose, A. Santini, A. Fortuna, M. L. García, A. M. Silva *et al.*, “Sugar-Lowering Drugs for Type 2 Diabetes Mellitus and Metabolic Syndrome—Review of Classical and New Compounds: Part-I,” *Pharmaceuticals*, vol. 12, no. 4, p. 152, 2019.

[103] National Institute for Health and Care Excellence, “NICE NG28. Type 2 diabetes in adults: management,” 2015.

[104] N. Hex, C. Bartlett, D. Wright, M. Taylor, and D. Varley, “Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costsand indirect societal and productivity costs,” *Diabetic Medicine*, vol. 29, no. 7, pp. 855–862, 2012.

[105] International Diabetes Federation, “IDF diabetes atlas, 10th edition,” 2021.

[106] M. J. Davies, D. A. D’Alessio, J. Fradkin, W. N. Kernan, C. Mathieu, G. Mingrone, P. Rossing, A. Tsapas, D. J. Wexler, and J. B. Buse, “Management of hyperglycemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD),” *Diabetes Care*, vol. 41, no. 12, pp. 2669–2701, 2018.

[107] S. D. Wiviott, I. Raz, M. P. Bonaca, O. Mosenzon, E. T. Kato, A. Cahn, M. G. Silverman, T. A. Zelniker, J. F. Kuder, S. A. Murphy *et al.*, “Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes,” *New England Journal of Medicine*, vol. 380, no. 4, pp. 347–357, 2019.

[108] H. J. Heerspink, B. V. Stefánsson, R. Correa-Rotter, G. M. Chertow, T. Greene, F.-F. Hou, J. F. Mann, J. J. McMurray, M. Lindberg, P. Rossing *et al.*, “Dapagliflozin in Patients with Chronic Kidney Disease,” *New England Journal of Medicine*, vol. 383, no. 15, pp. 1436–1446, 2020.

[109] K. Huynh, “Dapagliflozin—a breakthrough in the search for drugs to treat HFrEF,” *Nature Reviews Cardiology*, vol. 16, no. 12, pp. 700–700, 2019.

[110] P. McEwan, V. Foos, B. Martin, J. Chen, and M. Evans, “Estimating the value of sodium-glucose cotransporter-2 inhibitors within the context of contemporary guidelines and the totality of evidence,” *Diabetes, Obesity and Metabolism*, 2023.

[111] National Institute for Health and Care Excellence, “Type 2 diabetes in adults: management,” 2022. [Online]. Available: www.nice.org.uk/guidance/ng28

[112] C. C. Cowie, K. F. Rust, E. S. Ford, M. S. Eberhardt, D. D. Byrd-Holt, C. Li, D. E. Williams, E. W. Gregg, K. E. Bainbridge, S. H. Saydah *et al.*, “Full Accounting of Diabetes and Pre-Diabetes in the U.S. Population in 1988-1994 and 2005-2006,” *Diabetes Care*, vol. 32, no. 2, pp. 287–294, 2009.

[113] D. Bradley and W. Hsueh, “Type 2 Diabetes in the Elderly: Challenges in a Unique Patient Population,” *Journal of Geriatric Medicine and Gerontology*, vol. 2, no. 2, 2016.

[114] F. Zaccardi, D. R. Webb, T. Yates, and M. J. Davies, “Pathophysiology of type 1 and type 2 diabetes mellitus: a 90-year perspective,” *Postgraduate Medical Journal*, vol. 92, no. 1084, pp. 63–69, 2016.

[115] Diabetes UK, “How many people in the UK have diabetes?” 7 2023. [Online]. Available: https://www.diabetes.org.uk/professionals/position-statements-reports/statistics

[116] M. Roden and G. I. Shulman, “The integrative biology of type 2 diabetes,” *Nature*, vol. 576, no. 7785, pp. 51–60, 2019.

[117] U. Galicia-Garcia, A. Benito-Vicente, S. Jebari, A. Larrea-Sebal, H. Siddiqi, K. B. Uribe, H. Ostolaza, and C. Martín, “Pathophysiology of Type 2 Diabetes Mellitus,” *International Journal of Molecular Sciences*, vol. 21, no. 17, p. 6275, 2020.

[118] M. A. Atkinson, G. S. Eisenbarth, and A. W. Michels, “Type 1 diabetes,” *The Lancet*, vol. 383, no. 9911, pp. 69–82, 2014.

[119] S. Demir, P. P. Nawroth, S. Herzig, and B. Ekim Üstünel, “Emerging Targets in Type 2 Diabetes and Diabetic Complications,” *Advanced Science*, vol. 8, no. 18, p. 2100275, 2021.

[120] A. E. Butler and D. Misselbrook, “Distinguishing between type 1 and type 2 diabetes,” *BMJ*, vol. 370, 2020.

[121] World Health Organisation, “Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus,” 2011.

[122] A. Ramachandran, “Know the signs and symptoms of diabetes,” *The Indian Journal of Medical Research*, vol. 140, no. 5, p. 579, 2014.

[123] N. A. ElSayed, G. Aleppo, V. R. Aroda, R. R. Bannuru, F. M. Brown, D. Bruemmer, B. S. Collins, M. E. Hilliard, D. Isaacs, E. L. Johnson, S. Kahan, K. Khunti, J. Leon, S. K. Lyons, M. L. Perry, P. Prahalad, R. E. Pratley, J. J. Seley, R. C. Stanton, R. A. Gabbay, and on behalf of the American Diabetes Association, “6. Glycemic Targets: Standards of Care in Diabetes—2023,” *Diabetes Care*, vol. 46, pp. S97–S110, 12 2022. [Online]. Available: https://doi.org/10.2337/dc23-S006

[124] J. B. Buse, D. J. Wexler, A. Tsapas, P. Rossing, G. Mingrone, C. Mathieu, D. D'Alessio, and M. J. Davies, "2019 Update to: Management of Hyperglycemia in Type 2 Diabetes, 2018. A consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD)," *Diabetes Care*, vol. 43, pp. 487–493, 2020.

[125] K. Yau, A. Dharia, I. Alrowiyti, and D. Z. Cherney, "Prescribing SGLT2 Inhibitors in Patients With CKD: Expanding Indications and Practical Considerations," *Kidney International Reports*, vol. 7, no. 7, pp. 1463–1476, 2022.

[126] H. J. Curtis, J. M. Dennis, B. M. Shields, A. J. Walker, S. Bacon, A. T. Hattersley, A. G. Jones, and B. Goldacre, "Time trends and geographical variation in prescribing of drugs for diabetes in England from 1998 to 2017," *Diabetes, Obesity and Metabolism*, vol. 20, no. 9, pp. 2159–2168, 2018.

[127] J. M. Dennis, W. E. Henley, A. P. McGovern, A. J. Farmer, N. Sattar, R. R. Holman, E. R. Pearson, A. T. Hattersley, B. M. Shields, A. G. Jones, and on behalf of the MASTERMIND consortium, "Time trends in prescribing of type 2 diabetes drugs, glycaemic response and risk factors: a retrospective analysis of primary care data, 2010-2017," *Diabetes, Obesity and Metabolism*, vol. 21, no. 7, pp. 1576–1584, 2019.

[128] D. Soares, I. Palma, N. M. Helena, M. Fraga, J. Pereira, R. Guimaraes, C. M. Helena, A. Maia, and L. Ferreira, "Trends in Prescription of Anti-Diabetic Drugs and Metabolic Control in Type 2 Diabetes: 2015/2016 vs. 2020/2021," in *Endocrine Abstracts*, vol. 90. Bioscientifica, 2023.

[129] P. Bidulka, R. Mathur, D. G. Lugo-Palacios, S. O'Neill, A. Basu, R. J. Silverwood, P. Charlton, A. Briggs, L. Smeeth, A. I. Adler *et al.*, "Ethnic and socioeconomic disparities in initiation of second-line antidiabetic treatment for people with type 2 diabetes inEngland: A cross-sectional study," *Diabetes, Obesity and Metabolism*, vol. 25, no. 1, pp. 282–292, 2023.

[130] K. Khunti, S. Jabbour, X. Cos, S. Mudaliar, C. Mende, M. Bonaca, and P. Fioretto, "Sodium-glucose co-transporter-2 inhibitors in patients with type 2 diabetes: Barriers and solutions for improving uptake in routine clinical

practice," *Diabetes, Obesity and Metabolism*, vol. 24, no. 7, pp. 1187–1196, 2022.

[131] T. Hornick and D. C. Aron, "Managing diabetes in the elderly: Go easy, individualize," *Cleveland Clinic Journal of Medicine*, vol. 75, no. 1, p. 70, 2008.

[132] H. Sun, P. Saeedi, S. Karuranga, M. Pinkepank, K. Ogurtsova, B. B. Duncan, C. Stein, A. Basit, J. C. Chan, J. C. Mbanya *et al.*, "IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes Research and Clinical Practice*, vol. 183, p. 109119, 2022.

[133] R. Gómez-Huelgas, F. G. Peralta, L. R. Mañas, F. Formiga, M. P. Domingo, J. M. Bravo, C. Miranda, and J. Ena, "Treatment of type 2 diabetes mellitus in elderly patients," *Revista Clínica Española (English Edition)*, vol. 218, no. 2, pp. 74–88, 2018.

[134] A. D. Mooradian, "Mechanisms of Age-Related Endocrine Alterations: Part II," *Drugs & Aging*, vol. 3, pp. 131–146, 1993.

[135] G. Paolisso, "Pathophysiology of diabetes in elderly people." *Acta Biomedica: Atenei Parmensis*, vol. 81, pp. 47–53, 2010.

[136] S. J. Giddings, L. R. Carnaghi, and A. D. Mooradian, "Age-related changes in pancreatic islet cell gene expression," *Metabolism*, vol. 44, no. 3, pp. 320–324, 1995.

[137] S. B. Dybicz, S. Thompson, S. Molotsky, and B. Stuart, "Prevalence of Diabetes and the Burden of Comorbid Conditions Among Elderly Nursing Home Residents," *The American Journal of Geriatric pharmacotherapy*, vol. 9, no. 4, pp. 212–223, 2011.

[138] J. Freeman, "Management of hypoglycemia in older adults with type 2 diabetes," *Postgraduate Medicine*, vol. 131, no. 4, pp. 241–250, 2019.

[139] G. S. Meneilly and D. Tessier, "Diabetes in Elderly Adults," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 56, no. 1, pp. M5–M13, 2001.

[140] A. D. Mooradian, “Evidence-Based Management of Diabetes in Older Adults,” *Drugs & Aging*, vol. 35, pp. 1065–1078, 2018.

[141] S. Bellary, I. Kyrou, J. E. Brown, and C. J. Bailey, “Type 2 diabetes mellitus in older adults: clinical considerations and management,” *Nature Reviews Endocrinology*, vol. 17, no. 9, pp. 534–548, 2021.

[142] P. Bramlage, A. K. Gitt, C. Binz, M. Krekler, E. Deeg, and D. Tschöpe, “Oral antidiabetic treatment in type-2 diabetes in the elderly: balancing the need for glucose control and the risk of hypoglycemia,” *Cardiovascular Diabetology*, vol. 11, pp. 1–9, 2012.

[143] S. C. Johnston, “Combining ecological and individual variables to reduce confounding by indication:: Case study—subarachnoid hemorrhage treatment,” *Journal of Clinical Epidemiology*, vol. 53, no. 12, pp. 1236–1241, 2000.

[144] E. W. Gregg, E. Patorno, A. J. Karter, R. Mehta, E. S. Huang, M. White, C. J. Patel, A. T. McElvaine, W. T. Cefalu, J. Selby *et al.*, “Use of Real-World Data in Population Science to Improve the Prevention and Care of Diabetes-Related Outcomes,” *Diabetes Care*, vol. 46, no. 7, pp. 1316–1326, 2023.

[145] S. Vijan, D. Kent, and R. Hayward, “Are randomized controlled trials sufficient evidence to guide clinical practice in Type II (non-insulin-dependent) diabetes mellitus?” *Diabetologia*, vol. 43, pp. 125–130, 2000.

[146] R. E. Sherman, S. A. Anderson, G. J. Dal Pan, G. W. Gray, T. Gross, N. L. Hunter, L. LaVange, D. Marinac-Dabic, P. W. Marks, M. A. Robb *et al.*, “Real-World Evidence — What Is It and What Can It Tell Us?” *The New England Journal of Medicine*, vol. 375, no. 23, pp. 2293–2297, 2016.

[147] W. H. Herman, “Evidence-Based Diabetes Care,” *Clinical Diabetes*, vol. 20, no. 1, pp. 22–23, 2002.

[148] W. Yang, A. Zilov, P. Soewondo, O. M. Bech, F. Sekkal, and P. D. Home, “Observational studies: going beyond the boundaries of randomized controlled trials,” *Diabetes Research and Clinical Practice*, vol. 88, pp. S3–S9, 2010.

[149] N. Freemantle and T. Strack, “Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials,” *Journal of Clinical Epidemiology*, vol. 63, no. 10, pp. 1053–1058, 2010.

[150] A. T. N. Nair, A. Wesolowska-Andersen, C. Brorsson, A. L. Rajendrakumar, S. Hapca, S. Gan, A. Y. Dawed, L. A. Donnelly, R. McCrimmon, A. S. Doney *et al.*, “Heterogeneity in phenotype, disease progression and drug response in type 2 diabetes,” *Nature Medicine*, vol. 28, no. 5, pp. 982–988, 2022.

[151] C. Saunders, C. D. Byrne, B. Guthrie, R. Lindsay, J. McKnight, S. Philip, N. Sattar, J. Walker, S. Wild, and S. D. R. N. E. Group, “External validity of randomized controlled trials of glycaemic control and vascular disease: how representative are participants?” *Diabetic Medicine*, vol. 30, no. 3, pp. 300–308, 2013.

[152] B. Zinman, C. Wanner, J. M. Lachin, D. Fitchett, E. Bluhmki, S. Hantel, M. Mattheus, T. Devins, O. E. Johansen, H. J. Woerle *et al.*, “Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes,” *New England Journal of Medicine*, vol. 373, no. 22, pp. 2117–2128, 2015.

[153] A. McGovern, M. Feher, N. Munro, and S. de Lusignan, “Sodium-Glucose Co-transporter 2 (SGLT2) Inhibitor: Comparing Trial Data and Real-World Use,” *Diabetes Therapy*, vol. 8, pp. 365–376, 2017.

[154] A. Goldman, B. Fishman, G. Twig, E. Raschi, T. Cukierman-Yaffe, Y. Moshkovits, A. Pomerantz, I. Ben-Zvi, R. Dankner, and E. Maor, “The real-world safety profile of sodium-glucose co-transporter-2 inhibitors among older adults ($\geq$ 75 years): a retrospective, pharmacovigilance study,” *Cardiovascular Diabetology*, vol. 22, no. 1, p. 16, 2023.

[155] K. G. Young, E. Haider McInnes, R. J. Massey, A. R. Kahkoska, S. J. Pilla, S. Raghavan, M. A. Stanislawski, D. K. Tobias, A. P. McGovern, A. Y. Dawed *et al.*, “Precision medicine in type 2 diabetes: A systematic review of treatment effect heterogeneity for GLP1-receptor agonists and SGLT2-inhibitors,” *medRxiv*, pp. 2023–04, 2023.

[156] S. Schneeweiss and E. Patorno, “Conducting Real-world Evidence Stud-

ies on the Clinical Outcomes of Diabetes Treatments," *Endocrine Reviews*, vol. 42, no. 5, pp. 658–690, 2021.

[157] A. P. McGovern, M. Hogg, B. M. Shields, N. A. Sattar, R. R. Holman, E. R. Pearson, A. T. Hattersley, A. G. Jones, and J. M. Dennis, "Risk factors for genital infections in people initiating SGLT2 inhibitors and their impact on discontinuation," *BMJ Open Diabetes Research and Care*, vol. 8, no. 1, p. e001238, 2020.

[158] E. Patorno, A. Pawar, L. G. Bessette, D. H. Kim, C. Dave, R. J. Glynn, M. N. Munshi, S. Schneeweiss, D. J. Wexler, and S. C. Kim, "Comparative Effectiveness and Safety of Sodium–Glucose Cotransporter 2 Inhibitors Versus Glucagon-Like Peptide 1 Receptor Agonists in Older Adults," *Diabetes Care*, vol. 44, no. 3, pp. 826–835, 2021.

[159] Y.-M. Gao, S.-T. Feng, Y. Wen, T.-T. Tang, B. Wang, and B.-C. Liu, "Cardiorenal protection of SGLT2 inhibitors—Perspectives from metabolic reprogramming," *EBioMedicine*, vol. 83, 2022.

[160] A. Janež and P. Fioretto, "SGLT2 Inhibitors and the Clinical Implications of Associated Weight Loss in Type 2 Diabetes: A Narrative Review," *Diabetes Therapy*, vol. 12, pp. 2249–2261, 2021.

[161] L. C. Pinto, D. V. Rados, L. R. Remonti, C. K. Kramer, C. B. Leitao, and J. L. Gross, "Efficacy of SGLT2 inhibitors in glycemic control, weight loss and blood pressure reduction: a systematic review and meta-analysis," *Diabetology & Metabolic Syndrome*, vol. 7, no. 1, pp. 1–2, 2015.

[162] A. J. Scheen, "Efficacy/safety balance of DPP-4 inhibitors versus SGLT2 inhibitors in elderly patients with type 2 diabetes," *Diabetes & Metabolism*, vol. 47, no. 6, p. 101275, 2021.

[163] A. Tentolouris, P. Vlachakis, E. Tzeravini, I. Eleftheriadou, and N. Tentolouris, "SGLT2 Inhibitors: A Review of Their Antidiabetic and Cardioprotective Effects," *International Journal of Environmental Research and Public Health*, vol. 16, no. 16, p. 2965, 2019.

[164] B. C. Lupsa and S. E. Inzucchi, “Use of SGLT2 inhibitors in type 2 diabetes: weighing the risks and benefits,” *Diabetologia*, vol. 61, no. 10, pp. 2118–2125, 2018.

[165] A. J. Sinclair, B. Bode, S. Harris, U. Vijapurkar, W. Shaw, M. Desai, and G. Meininger, “Efficacy and Safety of Canagliflozin in Individuals Aged 75 and Older with Type 2 Diabetes Mellitus: A Pooled Analysis,” *Journal of the American Geriatrics Society*, vol. 64, no. 3, pp. 543–552, 2016.

[166] A. Avogaro, E. Delgado, and I. Lingvay, “When metformin is not enough: Pros and cons of SGLT2 and DPP-4 inhibitors as a second line therapy,” *Diabetes/metabolism research and reviews*, vol. 34, no. 4, p. e2981, 2018.

[167] A. Cove-Smith and M. Almond, “Management of urinary tract infections in the elderly,” *Trends in Urology, Gynaecology & Sexual Health*, vol. 12, no. 4, pp. 31–34, 2007.

[168] C. M. Apovian, J. Okemah, and P. M. O’Neil, “Body Weight Considerations in the Management of Type 2 Diabetes,” *Advances in Therapy*, vol. 36, pp. 44–58, 2019.

[169] O. R. Ghosh-Swaby, S. G. Goodman, L. A. Leiter, A. Cheng, K. A. Connelly, D. Fitchett, P. Jüni, M. E. Farkouh, and J. A. Udell, “Glucose-lowering drugs or strategies, atherosclerotic cardiovascular events, and heart failure in people with or at risk of type 2 diabetes: an updated systematic review and meta-analysis of randomised cardiovascular outcome trials,” *The Lancet Diabetes & Endocrinology*, vol. 8, no. 5, pp. 418–435, 2020.

[170] F. B. Ortega, C. J. Lavie, and S. N. Blair, “Obesity and Cardiovascular Disease,” *Circulation Research*, vol. 118, no. 11, pp. 1752–1770, 2016.

[171] K. Stenlöf, W. Cefalu, K.-A. Kim, M. Alba, K. Usiskin, C. Tong, W. Canovatchel, and G. Meininger, “Efficacy and safety of canagliflozin monotherapy in subjects with type 2 diabetes mellitus inadequately controlled with diet and exercise,” *Diabetes, Obesity and Metabolism*, vol. 15, no. 4, pp. 372–382, 2013.

[172] M. Roden, J. Weng, J. Eilbracht, B. Delafont, G. Kim, H. J. Woerle, and U. C. Broedl, “Empagliflozin monotherapy with sitagliptin as an active

comparator in patients with type 2 diabetes: a randomised, double-blind, placebo-controlled, phase 3 trial," *The Lancet Diabetes & Endocrinology*, vol. 1, no. 3, pp. 208–219, 2013.

[173] E. Ferrannini, S. J. Ramos, A. Salsali, W. Tang, and J. F. List, "Dapagliflozin Monotherapy in Type 2 Diabetic Patients With Inadequate Glycemic Control by Diet and Exercise: A randomized, double-blind, placebo-controlled, phase 3 trial," *Diabetes Care*, vol. 33, no. 10, pp. 2217–2224, 2010.

[174] S. G. Terra, K. Focht, M. Davies, J. Frias, G. Derosa, A. Darekar, G. Golm, J. Johnson, D. Saur, B. Lauring *et al.*, "Phase III, efficacy and safety study of ertugliflozin monotherapy in people with type 2 diabetes mellitus inadequately controlled with diet and exercise alone," *Diabetes, Obesity and Metabolism*, vol. 19, no. 5, pp. 721–728, 2017.

[175] B. Bode, K. Stenlöf, D. Sullivan, A. Fung, and K. Usiskin, "Efficacy and Safety of Canagliflozin Treatment in Older Subjects With Type 2 Diabetes Mellitus: A Randomized Trial," *Hospital Practice*, vol. 41, no. 2, pp. 72–84, 2013.

[176] B. Bode, K. Stenlöf, S. Harris, D. Sullivan, A. Fung, K. Usiskin, and G. Meininger, "Long-term efficacy and safety of canagliflozin over 104 weeks in patients aged 55–80 years with type 2 diabetes," *Diabetes, Obesity and Metabolism*, vol. 17, no. 3, pp. 294–303, 2015.

[177] S. Thayer, W. Chow, S. Korrer, and R. Aguilar, "Real-world evaluation of glycemic control among patients with type 2 diabetes mellitus treated with canagliflozin versus dipeptidyl peptidase-4 inhibitors," *Current Medical Research and Opinion*, vol. 32, no. 6, pp. 1087–1096, 2016.

[178] J. M. Dennis, K. G. Young, A. P. McGovern, B. A. Mateen, S. J. Vollmer, M. D. Simpson, W. E. Henley, R. R. Holman, N. Sattar, and E. R. Pearson, "Development of a treatment selection algorithm for SGLT2 and DPP-4 inhibitor therapies in people with type 2 diabetes: a retrospective cohort study," *The Lancet Digital Health*, vol. 4, pp. e873–e883, 2022.

[179] B. Neal, V. Perkovic, K. W. Mahaffey, D. De Zeeuw, G. Fulcher, N. Erondu, W. Shaw, G. Law, M. Desai, and D. R. Matthews, "Canagliflozin and Car-

diovascular and Renal Events in Type 2 Diabetes," *New England Journal of Medicine*, vol. 377, no. 7, pp. 644–657, 2017.

[180] J. Shikuma, R. Ito, J. Sasaki-Shima, A. Teshima, K. Hara, T. Takahashi, H. Sakai, T. Miwa, A. Kanazawa, and M. Odawara, "Changes in overactive bladder symptoms after sodium glucose cotransporter-2 inhibitor administration to patients with type 2 diabetes," *Practical Diabetes*, vol. 35, no. 2, pp. 47–50, 2018.

[181] J. Liu, L. Li, S. Li, P. Jia, K. Deng, W. Chen, and X. Sun, "Effects of SGLT2 inhibitors on UTIs and genital infections in type 2 diabetes mellitus: a systematic review and meta-analysis," *Scientific Reports*, vol. 7, no. 1, p. 2824, 2017.

[182] D. Vasilakou, T. Karagiannis, E. Athanasiadou, M. Mainou, A. Liakos, E. Bekiari, M. Sarigianni, D. R. Matthews, and A. Tsapas, "Sodium–Glucose Cotransporter 2 Inhibitors for Type 2 Diabetes: A Systematic Review and Meta Analysis," *Annals of Internal Medicine*, vol. 159, no. 4, pp. 262–274, 2013.

[183] K. Y. Thong, M. Yadagiri, D. J. Barnes, D. S. Morris, T. A. Chowdhury, L. L. Chuah, A. M. Robinson, S. C. Bain, K. A. Adamson, R. E. J. Ryder *et al.*, "Clinical risk factors predicting genital fungal infections with sodium–glucose cotransporter 2 inhibitor treatment: The ABCD nationwide dapagliflozin audit," *Primary Care Diabetes*, vol. 12, no. 1, pp. 45–50, 2018.

[184] P. Monteiro, R. M. Bergenstal, E. Toural, S. E. Inzucchi, B. Zinman, S. Hantel, S. G. Kiš, S. Kaspers, J. T. George, and D. Fitchett, "Efficacy and safety of empagliflozin in older patients in the EMPA-REG OUTCOME® trial," *Age and Ageing*, vol. 48, no. 6, pp. 859–866, 2019.

[185] O. Kinduryte Schorling, D. Clark, I. Zwiener, S. Kaspers, J. Lee, and H. Iliev, "Pooled Safety and Tolerability Analysis of Empagliflozin in Patients with Type 2 Diabetes Mellitus," *Advances in Therapy*, vol. 37, pp. 3463–3484, 2020.

[186] G. E. Umpierrez, M. B. Murphy, and A. E. Kitabchi, "Diabetic ketoacidosis

and hyperglycemic hyperosmolar syndrome," *Diabetes Spectrum*, vol. 15, no. 1, pp. 28–36, 2002.

[187] A. Puttanna and R. Padinjakara, "Diabetic ketoacidosis in type 2 diabetes mellitus," *Practical Diabetes*, vol. 31, no. 4, pp. 155–158, 2014.

[188] A. L. Peters, E. O. Buschur, J. B. Buse, P. Cohan, J. C. Diner, and I. B. Hirsch, "Euglycemic Diabetic Ketoacidosis: A Potential Complication of Treatment With Sodium–Glucose Cotransporter 2 Inhibition," *Diabetes Care*, vol. 38, no. 9, pp. 1687–1693, 2015.

[189] U.S. Food and Drug Administration, "FDA revises labels of SGLT2 inhibitors for diabetes to include warnings about," 2015, accessed on 07.07.2023. [Online]. Available: https://www.fda.gov/files/drugs/published/FDA-revises-labels-of-SGLT2-inhibitors-for-diabetes-to-include-warnings-about-too-much-acid-in-the-blood-and-serious-urinary-tract-infections.pdf

[190] European Medicines Agency, "EMA confirms recommendations to minimise ketoacidosis risk with SGLT2 inhibitors for diabetes," 2016, accessed on 07.07.2023. [Online]. Available: https://www.ema.europa.eu/en/medicines/human/referrals/sglt2-inhibitors

[191] M. Fralick, S. Schneeweiss, and E. Patorno, "Risk of Diabetic Ketoacidosis after Initiation of an SGLT2 inhibitor," *New England Journal of Medicine*, vol. 376, no. 23, pp. 2300–2302, 2017.

[192] P. Ueda, H. Svanström, M. Melbye, B. Eliasson, A.-M. Svensson, S. Franzén, S. Gudbjörnsdottir, K. Hveem, C. Jonasson, and B. Pasternak, "Sodium glucose cotransporter 2 inhibitors and risk of serious adverse events: nationwide register based cohort study," *BMJ*, vol. 363, 2018.

[193] Y. Chiba, Y. Kimbara, R. Kodera, Y. Tsuboi, K. Sato, Y. Tamura, S. Mori, H. Ito, and A. Araki, "Risk factors associated with falls in elderly patients with type 2 diabetes," *Journal of Diabetes and its Complications*, vol. 29, no. 7, pp. 898–902, 2015.

[194] T. Roman de Mettelinge, D. Cambier, P. Calders, N. Van Den Noortgate, and K. Delbaere, "Understanding the Relationship between Type 2 Dia-

betes Mellitus and Falls in Older Adults: A Prospective Cohort Study," *PloS ONE*, vol. 8, no. 6, p. e67055, 2013.

[195] A. J. Scheen, "Does lower limb amputation concern all SGLT2 inhibitors?" *Nature Reviews Endocrinology*, vol. 14, no. 6, pp. 326–328, 2018.

[196] U.S. Food and Drug Administration, "Drug and Safety Communications," 2017, accessed on 07.07.2023. [Online]. Available: https://www.fda.gov/downloads/Drugs/DrugSafety/UCM558427.pdf

[197] European Medicines Agency, "SGLT2 inhibitors: information on potential risk of toe amputation to be included in prescribing information," 2017. [Online]. Available: https://www.ema.europa.eu/en/medicines/human/referrals/sglt2-inhibitors-previously-canagliflozin

[198] J. Y. Yang, T. Wang, V. Pate, E. W. Gower, M. J. Crowley, J. B. Buse, and T. Stürmer, "Sodium-glucose co-transporter-2 inhibitor use and risk of lower-extremity amputation: Evolving questions, evolving answers," *Diabetes, Obesity and Metabolism*, vol. 21, no. 5, pp. 1223–1236, 2019.

[199] M. Fralick, S. C. Kim, S. Schneeweiss, B. M. Everett, R. J. Glynn, and E. Patorno, "Risk of amputation with canagliflozin across categories of age and cardiovascular risk in three US nationwide databases: cohort study," *BMJ*, vol. 370, 2020.

[200] E. Herrett, A. M. Gallagher, K. Bhaskaran, H. Forbes, R. Mathur, T. Van Staa, and L. Smeeth, "Data Resource Profile: Clinical Practice Research Datalink (CPRD)," *International Journal of Epidemiology*, vol. 44, no. 3, pp. 827–836, 2015.

[201] R. E. Ghosh, E. Crellin, S. Beatty, K. Donegan, P. Myles, and R. Williams, "How Clinical Practice Research Datalink data are used to support pharmacovigilance," *Therapeutic Advances in Drug Safety*, vol. 10, p. 2042098619854010, 2019.

[202] A. Wolf, D. Dedman, J. Campbell, H. Booth, D. Lunn, J. Chapman, and P. Myles, "Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum," *International Journal of Epidemiology*, vol. 48, no. 6, pp. 1740–1740g, 2019.

[203] L. R. Rodgers, M. N. Weedon, W. E. Henley, A. T. Hattersley, and B. M. Shields, “Cohort profile for the MASTERMIND study: using the Clinical Practice Research Datalink (CPRD) to investigate stratification of response to treatment in patients with type 2 diabetes,” *BMJ open*, vol. 7, no. 10, p. e017989, 2017.

[204] K. J. Jager, C. Zoccali, A. MacLeod, and F. W. Dekker, “Confounding: What it is and how to deal with it,” *Kidney International*, vol. 73, pp. 256–260, 2008.

[205] G. K. Dawwas, S. M. Smith, and H. Park, “Cardiovascular outcomes of sodium glucose cotransporter-2 inhibitors in patients with type 2 diabetes,” *Metabolism*, vol. 21, pp. 28–36, 2019.

[206] J. Bowden, B. Bornkamp, E. Glimm, and F. Bretz, “Connecting Instrumental Variable methods for causal inference to the Estimand Framework,” *Statistics in Medicine*, vol. 40, pp. 5605–5627, 2021.

[207] L. E. Pezzin, P. Laud, T. Yen, J. Neuner, and A. B. Nattinger, “Re-examining the Relationship of Breast Cancer Hospital and Surgical Volume to Mortality: An Instrumental Variable Analysis,” *Medical Care*, vol. 53, no. 12, p. 1033, 2015.

[208] G. D. Smith and S. Ebrahim, ““Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease?” *International Journal of Epidemiology*, vol. 32, pp. 1–22, 2003.

[209] M. G. Weiner, D. Xie, and R. L. Tannen, “Replication of the Scandinavian Simvastatin SurvivalStudy using a primary care medical record database-prompted exploration of a new method to addressunmeasured confounding,” *Pharmacoepidemiology and Drug Safety*, vol. 17, pp. 661–670, 2008.

[210] A. Gallagher, F. de Vries, and T. van Staa, “Prior event rate ratio adjustment: a magic bullet or more of the same?” vol. 18. JOHN WILEY & SONS LTD THE ATRIUM, SOUTHERN GATE, CHICHESTER PO19 8SQ, W, 2009, pp. S14–S15.

[211] J. L. Bernal, S. Cummins, and A. Gasparrini, “Difference in difference, controlled interrupted time series and synthetic controls,” *International Journal of Epidemiology*, vol. 48, pp. 2062–2063, 2019.

[212] P. Craig, C. Cooper, D. Gunnell, S. Haw, K. Lawson, S. Macintyre, D. Ogilvie, M. Petticrew, B. Reeves, and M. Sutton, “Using natural experiments to evaluate population health interventions: new Medical Research Council guidance,” *Journal of Epidemiol Community Health*, vol. 66, pp. 1182–1186, 2012.

[213] P. C. Rockers, J.-A. Røttingen, I. Shemilt, P. Tugwell, and T. Bärnighausen, “Inclusion of quasi-experimental studies in systematic reviews of health systems research,” *Health Policy*, vol. 119, pp. 511–521, 2015.

[214] S. B. Soumerai, D. Starr, and S. R. Majumdar, “How Do You Know Which Health Care Effectiveness Research You Can Trust? A Guide to Study Design for the Perplexed,” *Preventing Chronic Disease*, vol. 12, 2015.

[215] T. J. Leeper, *margins: Marginal Effects for Model Objects*, 2021, r package version 0.3.26.

[216] A. Huitfeldt, M. J. Stensrud, and E. Suzuki, “On the collapsibility of measures of effect in the counterfactual causal framework,” *Emerging Themes in Epidemiology*, vol. 16, pp. 1–5, 2019.

[217] M. Slotani, “Tolerance regions for a multivariate normal population,” *Annals of the Institute of Statistical Mathematics*, vol. 16, pp. 135–153, 1964.

[218] W. G. Cochran, “The Combination of Estimates from Different Experiments,” *Biometrics*, vol. 10, pp. 101–129, 1954.

[219] T. P. Morris, I. R. White, and M. J. Crowther, “Using simulation studies to evaluate statistical methods,” *Statistics in Medicine*, vol. 38, pp. 2074–2102, 5 2019.

[220] D. Small, “Mediation Analysis Without Sequential Ignorability: Using Baseline Covariates Interacted with Random Assignment as Instrumental Variables,” *Journal of Statistical Research*, vol. 46, pp. 91–103, 2012.

[221] W. Spiller, W. Slichter, J. Bowden, and D. S. G., “Detecting and correcting for bias in Mendelian randomization analyses using Gene-by-Environment interactions,” *International Journal of Epidemiology*, vol. 48, pp. 702–712, 2019.

[222] National Institute for Health and Care Excellence, “Type 2 diabetes in adults: management,” 2020. [Online]. Available: www.nice.org.uk/guidance/ng28

[223] O. Montvida, J. Shaw, J. A. John, F. Stringer, and S. K. P. Paul, “Long-term Trends in Antidiabetes Drug Usage in the US: Real-world Evidence in Patients Newly Diagnosed with Type 2 Diabetes,” *Diabetes Care*, vol. 41, pp. 69–78, 2018.

[224] J. M. Dennis, “Precision Medicine in Type 2 Diabetes: Using Individualized Prediction Models to Optimize Selection of Treatment,” *Diabetes*, vol. 69, no. 10, pp. 2075–2085, 2020.

[225] C. M. Booth and I. F. Tannock, “Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence,” *British Journal of Cancer*, vol. 110, pp. 551–555, 2014.

[226] W. Hinton, M. Feher, N. Munro, M. Walker, and S. de Lusignan, “Real-world prevalence of the inclusion criteria for the LEADER trial: Data from a national general practice network,” *Diabetes, Obesity and Metabolism*, vol. 21, pp. 1661–1667, 7 2019.

[227] D. E. Ho, K. Imai, G. King, and E. A. Stuart, “MatchIt: Nonparametric Preprocessing for Parametric Causal Inference,” *Journal of Statistical Software*, vol. 42, no. 8, pp. 1–28, 2011.

[228] J. M. Dennis, K. G. Young, A. P. McGovern, B. A. Mateen, S. J. Vollmer, M. D. Simpson, W. E. Henley, R. R. Holman, N. Sattar, and E. R. Pearson, “Derivation and validation of a type 2 diabetes treatment selection algorithm for SGLT2-inhibitor and DPP4-inhibitor therapies based on glucose-lowering efficacy: cohort study using trial and routine clinical data,” *medRxiv*, 2021.

[229] T. Ye, A. Ertefaie, J. Flory, H. Sean, and S. D. Small, “Instrumented Difference-in-Differences,” *arXiv:2011.03593 [stat.ME]*, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2011.03593

[230] F. Yang, S. A. Lorch, and D. S. Small, “ESTIMATION OF CAUSAL EFFECTS USING INSTRUMENTAL VARIABLES WITH NONIGNORABLE MISSING COVARIATES: APPLICATION TO EFFECT OF TYPE OF DELIVERY NICU ON PREMATURE INFANTS,” *The Annals of Applied Statistics*, pp. 48–73, 2014.

[231] M. J. Uddin, R. H. H. Groenwold, A. de Boer, A. S. M. Afonso, P. Primatesta, C. Becker, S. V. Belitser, A. W. Hoes, K. C. B. Roes, and O. H. Klungel, “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*, vol. 25, pp. 132–141, 2016.

[232] I. G. G. Kreft, “Are multilevel techniques necessary? An overview, including simulation studies,” *Unpublished manuscript, California State University, Los Angeles*, 1996.

[233] T. A. B. Snijders and R. J. Bosker, *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. SAGE, 2011.

[234] J. Hook and R. van de Schoot, *Multilevel Analysis: Techniques and Applications*. Routledge, 2010.

[235] M. Sharma, I. Nazareth, and I. Petersen, “Trends in incidence, prevalence and prescribing in type 2 diabetes mellitus between 2000 and 2013 in primary care: a retrospective cohort study,” *BMJ open*, vol. 6, p. e010210, 2016.

[236] A. J. Scheen, “SGLT2 versus DPP4 inhibitors for type 2 diabetes,” *The Lancet Diabetes & Endocrinology*, vol. 1, no. 3, pp. 168–170, 2013.

[237] H. Wang, R. L. M. Cordiner, Y. Huang, L. Donnelly, S. Hapca, A. Collier, J. McKnight, B. Kennon, F. Gibb, and P. McKeigue, “Cardiovascular Safety in Type 2 Diabetes With Sulfonylureas as Second-line Drugs: A Nationwide

Population-Based Comparative Safety Study," *Diabetes Care*, p. dc221238, 2023.

[238] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, pp. 427–437, 2009.

[239] N. A. ElSayed, G. Aleppo, V. R. Aroda, R. R. Bannuru, F. M. Brown, D. Bruemmer, B. S. Collins, M. E. Hilliard, D. Isaacs, E. L. Johnson *et al.*, "1. Improving Care and Promoting Health in Populations: Standards of Care in Diabetes—2023," *Diabetes Care*, vol. 46, no. Supplement_1, pp. S10–S18, 2023.

[240] E. Brown, S. P. Rajeev, D. J. Cuthbertson, and J. P. Wilding, "A review of the mechanism of action, metabolic profile and haemodynamic effects of sodium-glucose co-transporter-2 inhibitors," *Diabetes, Obesity and Metabolism*, vol. 21, pp. 9–18, 2019.

[241] H. Storgaard, L. L. Gluud, C. Bennett, M. F. Grøndahl, M. B. Christensen, F. K. Knop, and T. Vilsbøll, "Benefits and Harms of Sodium-Glucose Co-Transporter 2 Inhibitors in Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis," *PloS ONE*, vol. 11, no. 11, p. e0166125, 2016.

[242] F. Zaccardi, D. R. Webb, Z. Z. Htike, D. Youssef, K. Khunti, and M. J. Davies, "Efficacy and safety of sodium-glucose co-transporter-2 inhibitors in type 2 diabetes mellitus: systematic review and network meta-analysis," *Diabetes, Obesity and Metabolism*, vol. 18, no. 8, pp. 783–794, 2016.

[243] A. Sinclair, B. Bode, S. Harris, U. Vijapurkar, C. Mayer, A. Fung, W. Shaw, K. Usiskin, M. Desai, and G. Meininger, "Efficacy and safety of canagliflozin compared with placebo in older patients with type 2 diabetes mellitus: a pooled analysis of clinical studies," *BMC Endocrine Disorders*, vol. 14, pp. 1–11, 2014.

[244] J. Pearson-Stuttard, Y. J. Cheng, J. Bennett, E. P. Vamos, B. Zhou, J. Valabhji, A. J. Cross, M. Ezzati, and E. W. Gregg, "Trends in leading causes of hospitalisation of adults with diabetes in England from 2003 to 2018: an

epidemiological analysis of linked primary care records," *The Lancet Diabetes & Endocrinology*, vol. 10, no. 1, pp. 46–57, 2022.

[245] M. Baiocchi, J. Cheng, and D. S. Small, "Instrumental variable methods for causal inference," *Statistics in Medicine*, vol. 33, no. 13, pp. 2297–2340, 2014.

[246] N. M. Davies, G. D. Smith, F. Windmeijer, and R. M. Martin, "COX-2 Selective Nonsteroidal Anti-inflammatory Drugs and Risk of Gastrointestinal Tract Complications and Myocardial Infarction: An Instrumental Variable Analysis," *Epidemiology*, pp. 352–362, 2013.

[247] J. Parkkari, P. Kannus, M. Palvanen, A. Natri, J. Vainio, H. Aho, I. Vuori, and M. Järvinen, "Majority of Hip Fractures Occur as a Result of a Fall and Impact on the Greater Trochanter of the Femur: A Prospective Controlled Hip Fracture Study with 206 Consecutive Patients," *Calcified Tissue International*, vol. 65, pp. 183–187, 1999.

[248] G. Schernthaner, J. L. Gross, J. Rosenstock, M. Guarisco, M. Fu, J. Yee, M. Kawaguchi, W. Canovatchel, and G. Meininger, "Canagliflozin Compared With Sitagliptin for Patients With Type 2 Diabetes Who Do Not Have Adequate Glycemic Control With Metformin Plus Sulfonylurea," *Diabetes Care*, vol. 36, no. 9, pp. 2508–2515, 2013.

[249] F. Lavalle-González, A. Januszewicz, J. Davidson, C. Tong, R. Qiu, W. Canovatchel, and G. Meininger, "Efficacy and safety of canagliflozin compared with placebo and sitagliptin in patients with type 2 diabetes on background metformin monotherapy: a randomised trial," *Diabetologia*, vol. 56, pp. 2582–2592, 2013.

[250] G. Schernthaner, F. J. Lavalle-González, J. A. Davidson, H. Jodon, U. Vijapurkar, R. Qiu, and W. Canovatchel, "Canagliflozin provides greater attainment of both HbA1c and body weight reduction versus sitagliptin in patients with type 2 diabetes," *Postgraduate Medicine*, vol. 128, no. 8, pp. 725–730, 2016.

[251] P. S. Wang, S. Schneeweiss, J. Avorn, M. A. Fischer, H. Mogun, D. H. Solomon, and M. A. Brookhart, "Risk of Death in Elderly Users of Con-

ventional vs. Atypical Antipsychotic Medications," *New England Journal of Medicine*, vol. 353, pp. 2335–2341, 2005.

[252] Y. Xie, B. Bowe, H. Xian, T. Loux, J. B. McGill, and Z. Al-Aly, "Comparative effectiveness of SGLT2 inhibitors, GLP-1 receptor agonists, DPP-4 inhibitors, and sulfonylureas on risk of major adverse cardiovascular events: emulation of a randomised target trial using electronic health records," *The Lancet Diabetes & Endocrinology*, 2023.

[253] N. Abate and M. Chandalia, "The impact of ethnicity on type 2 diabetes," *Journal of Diabetes and its Complications*, vol. 17, no. 1, pp. 39–58, 2003.