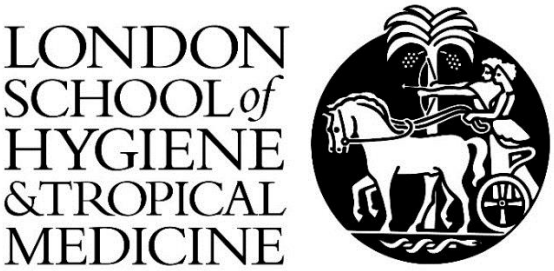


Application of Marginal Structural Models with Inverse Probability of Treatment Weighting in Electronic Health Records to Investigate the Benefits and Risks of First Line Type II Diabetes Treatments.

RUTH ESTHER FARMER

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of London

JUNE 2017

Department of Non Communicable Diseases Epidemiology

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE

Funded by the Medical Research Council London Hub for Trials Methodology Research

Research group affiliation: Electronic Health Records Research group, LSHTM.

DECLARATION

I Ruth Farmer, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: Date:

Use of own published material:

A version of the work presented in chapter 2 of this thesis was published prior to submission as a systematic review in the International Journal of Epidemiology, doi: 10.1093/ije/dyw275. The reference is given at the beginning of the chapter, and the published pdf version is presented as an appendix.

ABSTRACT

Background:

Electronic healthcare records (EHRs) provide opportunities to estimate the effects of type two diabetes (T2DM) treatments on outcomes such as cancer and cardiovascular disease. Marginal structural models (MSMs) with inverse probability of treatment weights (IPTW) can correctly estimate the causal effect of time-varying treatment in the presence of time-dependent confounders such as HbA1c. Dynamic MSMs can be used to compare dynamic treatment strategies. This thesis applies weighted MSMs and dynamic MSMs to explore risks and benefits of early-stage T2DM treatments, and considers the practicalities/impact of using these models in a complex clinical setting with a challenging data source.

Methods and Findings:

A cohort of patients with newly diagnosed T2DM was identified from the Clinical Practice Research Datalink. MSMs with IPTW were used to estimate the causal effect of metformin monotherapy on cancer risk, and the effects of metformin and sulfonylurea monotherapies on risks of MI, stroke, all-cause mortality, and HbA1c trajectory. Dynamic MSMs were implemented to compare HbA1c thresholds for treatment initiation on risks of MI, stroke, all-cause mortality (ACM) and glucose control. No association was found between metformin use and cancer risk. Metformin and sulfonylureas led to better HbA1c control than diet only, as expected, and there was some evidence of reduced MI risk with long-term metformin use. Changes in estimates between standard models and weighted models were generally in the expected direction given hypothesised time-dependent confounding. For stroke and ACM, results were less conclusive, with some suggestions of residual confounding. Higher HbA1c thresholds for treatment initiation reduced the likelihood of reaching target HbA1c, and there was a suggestion that higher initiation thresholds increased MI risk.

Conclusions:

Fitting weighted MSMs and dynamic MSMs was feasible using routine primary care data. The models appeared to work well in controlling for strong time-dependent confounding with short-term outcomes; results for longer-term outcomes were less conclusive.

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisors Dr. Krishnan Bhaskaran and Dr. Deborah Ford for their guidance. In addition, I am grateful to my advisory committee, Prof Liam Smeeth, Prof. Nishi Chaturvedi and Prof. Rick Kaplan for their valuable input, and also to other members of the electronic health records team at LSHTM for their helpful advice.

I would also like to acknowledge my funders, the MRC London Hub for Trials Methodology Research

Finally, I wish to thank my family and friends, especially my father for his encouragement, and James, who has given me so much support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 TYPE 2 DIABETES

1.1.1 Prevalence, risk factors and economic burden.

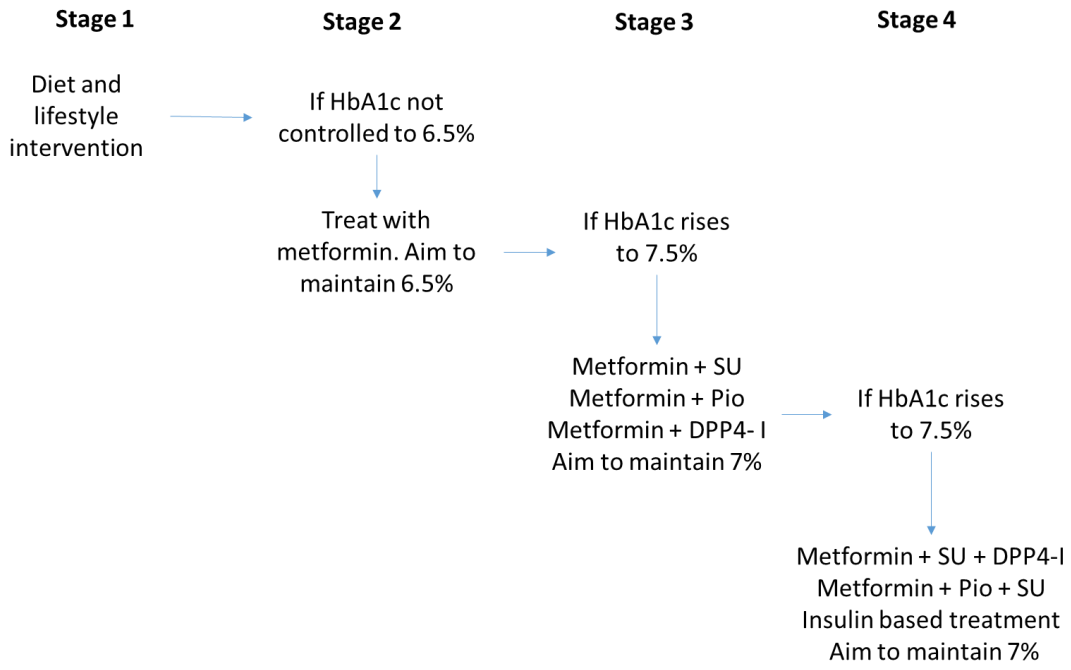
Type 2 Diabetes Mellitus (T2DM) is a condition related to a person's ability to produce and process insulin. Sufferers have difficulty in controlling blood glucose levels [1]. T2DM affects approximately 5.5% of the UK population [2], and this number is expected to rise to just under 10% by 2030 [3]. Key risk factors for T2DM are a family history , being overweight, older age, history of high blood pressure, and ethnicity (e.g. South Asian, African and African -Caribbean ethnicities have an elevated risk) [4]. It has also been suggested that smoking may be an independent risk factor for T2DM [5], possibly as a result of its effect on insulin resistance [6]. The strongest risk factors however, are those associated with being overweight. Specifically, a large waist circumference and weight gain in early adulthood have both been shown to increase risk [7, 8]. The American Diabetes Association estimates that after adjustment for age, gender and ethnicity, the average medical cost of a patient with diabetes is 2.3 times what it would be in the absence of diabetes, with an overall cost for diabetes in the USA of $245 billion in 2012 [9]. In the UK, the cost to the National Health Service (NHS) for diabetes is expected to be between £13 billion and £20 billion by 2035, which will be approximately 17% of overall expenditure [10]. Around 80% of this is attributed to treating complications (see 1.1.3) [11]. Since T2DM accounts for around 90% of diabetes cases [2], it is clear that T2DM is a large and increasing economic burden worldwide.

1.1.2 Management of T2DM

T2DM is treated with diet and exercise and/or anti-diabetes drugs, and is predominantly managed within primary care [12]. Figure 1.1 shows a simplified version of the recommended treatment pathways defined by the National Institute for Health and Care Excellence (NICE) [13]. For initial control of T2DM, the most common pharmacological treatment is the biguanide metformin [14]. This is a widely available drug that is also indicated for polycystic ovary syndrome [15]. Use of this as the first line treatment for T2DM has increased rapidly since the year 2000 [16]. As an alternative first line therapy, patients may be prescribed sulfonylureas, though use of these has declined in recent years [16] due to guidelines changing to advocate

metformin as the preferred first line therapy. In later stages treatment is usually intensified by adding additional oral anti-diabetes drugs (OADs) if glucose control is not well maintained with monotherapy. In newly diagnosed T2DM the current UK guidelines advocate the initiation of metformin if diet and lifestyle interventions cannot control blood glucose levels (as measured by glycated haemoglobin (HbA1c)) to below 6.5%. Other factors also have an influence on treatment decisions, for example, sulfonylureas are less likely to be used in overweight patients, and metformin is contra-indicated in patients with renal failure [15].

Figure 1.1 Simplified diagram of care pathway for type 2 diabetes in the UK, as developed by NICE and presented in the 2015 updated guidelines [13]



SU Sulfonylureas **Pio** Pioglitazone **DPP4-I** Dipeptidyl peptidase 4 inhibitors

1.1.3 Complications of T2DM and associations with anti-diabetes drugs

There is evidence that T2DM conveys additional risks of many other complications, for example, retinopathy, neuropathy, kidney disease, foot problems, cardiovascular disease (CVD) and cancer [2, 17-20]. As such, the potential effects of diabetes therapy on these excess risks are important to understand. A recent example of a potentially important treatment effect is the possible protective effect of metformin on risk of cancer [21-27]. In the last decade, multiple observational studies have suggested a reduced risk of both cancer incidence and mortality in

T2DM patients using metformin. A full systematic review of the evidence for this will be presented in chapter 2. As brief illustrations, studies by both Currie et al. [26] and Libby et al. [27] , using primary care data from the UK, estimated a decreased risk of cancer among metformin users of 27% and 37% respectively. However, other studies found no evidence of a protective effect, such as a study by Tsilidis [28] that found a HR for metformin use of 0.96 (0.89-1.04). In other contexts, concerns have been raised over diabetes medications. Multiple observational studies have suggested an increased risk of cardiovascular and all-cause mortality with use of sulfonylureas, as shown by a systematic review in 2013 [29]. However, the authors did acknowledge strong heterogeneity in their findings and potential for bias, and a recent meta-analyses of randomised studies did not find the same increased risk [30]. Long term clinical trials, in particular the UK Prospective Diabetes Study [31] have also suggested a protective effect of metformin on risk of myocardial infarction (MI), stroke and all- cause mortality in patients with T2DM compared to a lifestyle intervention. This may explain the apparent harmful effect of sulfonylureas, since many observational studies use metformin monotherapy as the comparator agent [32, 33]. However, even the beneficial effect of metformin is not confirmed. For example, a meta-analysis and meta-regression of randomised trials found that effect of metformin on a range of cardiovascular outcomes appeared to be dependent upon the age of the included patients and on the comparator group [34].

1.2 TIME-DEPENDENT CONFOUNDING

A major issue in studying the risks and benefits of early stage type 2 diabetes treatments, is that many risk factors for the complications outlined above, such as BMI and glucose control are themselves associated with diabetes severity. As such, they are predictors for initiation of OADs through time. In other words, at any given time, patients initiating OADs are likely to have an underlying risk of complications that is different to patients continuing a lifestyle and diet intervention. This can be difficult to control for, particularly when modelling a time-varying treatment. Specifically, since OADs are prescribed to control diabetes severity by way of glucose control, it is clear that diabetes severity is both a confounder of the association between treatment and outcome, but also on the causal pathway between treatment and outcome. In this situation, difficulties arise in estimating the overall effect of treatment on outcome using standard statistical methods.

Any time-dependent covariate L that is predictive of an outcome Y, and also of a time-varying exposure of interest X (often a treatment) is known as a time-dependent confounder. If past values of the treatment X also influence future values of L, then this confounder is referred to as a “time-dependent confounder affected by prior treatment”. This key definition is outlined in

Box 1.1 with an accompanying diagram. For clarity, we assume that the associations between L and X and L and Y are not both results of unmeasured common causes, as in this case L is not a confounder and treating it as such would induce selection bias [35, 36].

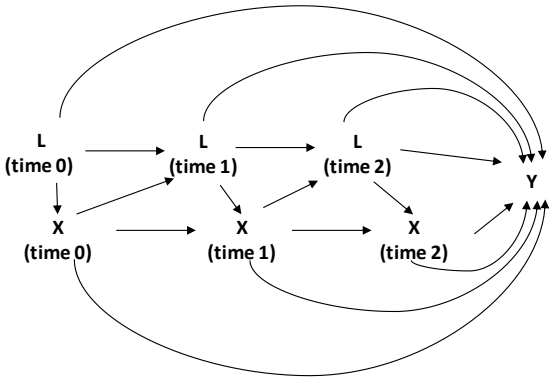
A variable is a **time-dependent (or time-varying) confounder affected by prior treatment** if it satisfies the following conditions:

1. The variable changes through time.

2. The variable is **predictive of** changes in **treatment**.

3. The variable is also **associated with the outcome** of interest.

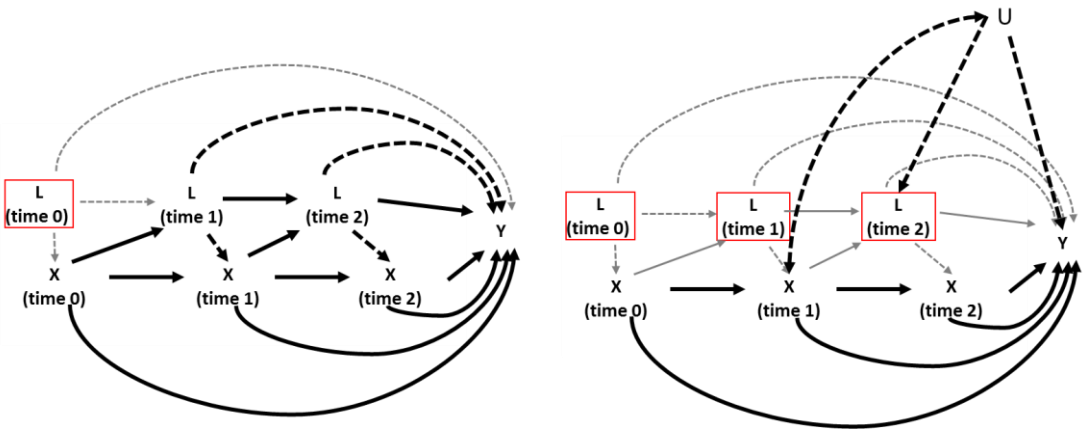
4. The variable **is also affected by past treatment.**



Box 1.1: Definition of a time-dependent confounder affected by prior treatment

In this situation standard statistical methods are unable to estimate the overall effect of treatment on outcome. To explain why, we consider the two options available for covariate adjustment. Firstly, an adjustment could be made for the value of L at time 0 (Figure 1.2, left). This would control for the initial confounding , however when exposure is time updated at time 1 and 2, the baseline value of L may no longer reflect the risk of the outcome Y at that time, and so residual confounding will remain (as indicated by the bold dashed lines). Secondly, the adjustment for L could be time updated (Figure 1.2, right). This will remove any residual confounding, but will also remove any effect of treatment that acts via L and so the overall effect of treatment cannot be established. Further, conditioning on time updated values of L may cause further confounding if there are risk factors for outcome and L that have not been accounted for, by inducing an association between X and U [36] as shown in the right hand pane of Figure 1.2.

Figure 1.2 Visualisation of the issues with standard methods of adjustment in the presence of time-dependent confounders affected by prior treatment.



Left – Baseline adjustment only. Right – time updated adjustment
Box indicates adjustment. Bold lines indicate those included in the estimate of treatment effect under the different options. Dotted lines indicate confounding pathways. Solid lines indicate assumed causal pathways.

1.3 THESIS MOTIVATION AND AIMS

Causal methodology has been developed to allow the correct estimation of the effect of time-varying treatment on outcome in the presence of time-dependent confounders [37-39]. One method in particular, the method of inverse probability of treatment weighting (IPTW) (sometimes more generally referred to as inverse probability weighting (IPW)) of marginal structural models (MSMs), is relatively intuitive and has been used successfully in other disease areas, such as establishing the causal effect of antiretroviral therapy on survival in HIV in the presence of time-dependent confounding by CD4 count [40, 41].

There are relatively few applications of IPTW of MSMs in the diabetes context, and an anticipated issue is that well-defined treatment guidelines may cause issues with the positivity assumption (necessary for MSMs, see 4.3.4.5), where certain treatment patterns are entirely determined by a particular covariate history. As described in 5.1, the data source for this thesis will be a large database of routinely collected primary care data. Such data sources are a valuable resource, allowing pharmacoepidemiological questions to be studied in large population-based samples. Although applications of MSMs in routinely collected electronic healthcare records (EHR) exist [42], some aspects of EHR data may cause uncertainty around the necessary assumptions. For example, there may be concerns of unmeasured or residual confounding from risk factors that are either not recorded, or are recorded with low accuracy. Additionally, frequency of visits to the GP will affect the opportunity to receive treatment, the recording of key information, and could be associated with underlying health.

1.3.1 Thesis aims

The main aims of this thesis are:

1. **To investigate, via the use of inverse probability of treatment weighting of MSMs, the potential risks and benefits of early stage type 2 diabetes treatments, focussing on areas of current interest in diabetes epidemiology which are likely to be affected by time-dependent confounding.**

Within this, the specific objectives are:

a) To understand how the question of whether metformin affects cancer risk has been previously addressed in terms of study designs, data sources, statistical methodology and risk of various biases; including how issues of time-dependent confounding may have been avoided/addressed.

b) To use IPTW of MSMs to estimate the effect of metformin monotherapy vs lifestyle intervention only on risk of cancer in a cohort patients with newly diagnosed T2DM.

c) To use IPTW of MSMs to estimate effect of metformin and sulfonylureas vs lifestyle intervention only on risk of cardiovascular events, all-cause mortality and glucose control in a cohort patients with newly diagnosed T2DM.

d) To use existing extensions to the methodology to compare risk of cardiovascular events, all-cause mortality and glucose control between "dynamic" treatment strategies. For example, the risk could be compared for "treat with metformin when HbA1c rises above 6.5%" vs ""treat with metformin when HbA1c rises above 8%".

2. To investigate the practicalities and impact of using inverse probability of treatment weighting of MSMs to adjust for anticipated time-dependent confounding in a complex clinical setting with a challenging data source.

Explicitly, this aim has two key objectives as follows:

e) To descriptively examine whether both the anticipated issues with positivity in the diabetes context, and potential issues of visit frequency appear to be present in the data.

f) To compare the estimated effects of treatment obtained from the MSMs to those obtained via standard analysis methods. Specifically, the differences between methods will be compared in situations where the true effect of the drugs of interest are both unknown (cancer, CV outcomes, mortality) and better established (glucose control).

1.4 THESIS OUTLINE

Chapter two presents a full systematic review and bias evaluation of the current literature on metformin and cancer risk, with a thorough discussion of the advantages and limitations of differing methodological approaches that have been used. Chapter three then presents a further short review of the existing literature on the risks and benefits of treatment strategies using metformin and sulfonylureas on cardiovascular outcomes and all-cause mortality.

Following this in chapter four; the concepts of causal inference are introduced, and the method of inverse probability of treatment weighting of MSMs described. Details of the data source will then be introduced in chapter five, along with a description of how the cohort of patients with T2DM used within this thesis were identified.

To establish the feasibility of the method, and to gain insight into the extent to which anticipated issues of positivity and visit frequency may be present, chapter six presents some initial descriptive analyses. Specifically, treatment patterns within the identified cohort are examined, and the measurement frequencies of key time-dependent confounders described.

In chapters seven, eight and nine; the three main pieces of analysis are presented. Chapter seven presents a study investigating the effect of metformin on risk of cancer incidence. In chapter eight, the effect of metformin and sulfonylureas on risk of MI, stroke, all-cause mortality, and longitudinal HbA1c control are estimated. In chapter nine, treatment strategies of the form "treat when HbA1c raises above x%" (known as dynamic treatments), will be compared for varying values of x. The aim of this will be to investigate whether the current HbA1c threshold of 6.5% for treatment initiation is appropriate in terms of maintaining good glucose control and minimising risk of cardiovascular events and mortality.

In the final chapter, the overall conclusions, and the key observed limitations of the methodology are discussed. Ideas for further work are also presented.

1.5 ETHICAL APPROVALS

The necessary ISAC approval and ethical approval from the London School of Hygiene and Tropical Medicine (LSHTM) was obtained (see appendices 1 and 2 for LSHTM Ethics and ISAC approval and associated protocol respectively).

1.6 ORIGINAL CONTRIBUTION

This thesis contributes original research to the existing field of diabetes pharmacoepidemiology in several ways. Firstly, although other systematic reviews of the effect of metformin on cancer exist; the review presented in chapter 2 is the first (as far as could be identified) that systematically summarises the existing research both in terms of estimated effects, and a comprehensive objective bias evaluation, to identify which existing studies appear the most reliable. Secondly, no existing literature could be found that applies a marginal structural model approach to any of the associations examined within this thesis. Approaching these questions from a different perspective, using an approach that allows (in theory) for the correct estimation of the effect of a time-varying treatment in the presence of time-dependent confounders affected by prior treatment, could potentially add to our understanding of the causal effect of first line diabetes treatments on important clinical outcomes. Thirdly, the extension of MSMs to the dynamic treatment setting has not yet been applied to examine strategies for first line diabetes interventions. This work could therefore provide insight into whether the current guidelines for initiation of first line therapy are optimal to reduce risk of later complications. Finally, from a methodological perspective, the experience of applying MSMs in a complex setting using large scale EHRs may be informative for future applications of the methodology in the growing field of e-health research.

2 A SYSTEMATIC REVIEW OF METFORMIN AND RISK OF CANCER IN TYPE 2 DIABETES

2.1 RATIONALE AND AIM

The aim of this review was to summarise existing observational studies investigating possible associations between metformin use and cancer risk in patients with T2DM, and to systematically examine the research design, analysis methods, and risks of bias. In line with the overall aims of the thesis, one of the key issues to examine was the extent to which the existing studies could be affected by time-dependent confounding affected by prior treatment. A secondary aim was to use meta-regression to investigate whether it was possible to quantify the extent to which study design characteristics and possible biases may account for the differences between study estimates. A version of this review was published in the International Journal of Epidemiology prior to submission of this thesis [43], and the published pdf is presented in appendix 3.

2.2 METHODS

2.2.1 Search strategy

MEDLINE was searched using OvidSP on 30th May 2014 for all English language articles on cancer risk and type 2 diabetes treatments from 1946 onwards. The search involved using MeSH headings as well as key word searches in the title and abstract. The full search terms are presented in Box 2.1. Conference abstracts and unpublished studies were excluded.

2.2.2 Screening strategy

Articles were included in the review if they were of a standard epidemiological design (i.e. cohort study or case control study) and presented original observational research. Reviews and meta-analyses were not included. Studies were required to present a measure of estimated effect of metformin on risk of cancer incidence (either all cancer or site specific) in patients with T2DM, with age adjustment as a minimum. Studies restricted to populations with additional co-morbidities or diseases (other than diabetes) were excluded.

1. Metformin/ or Insulin, Isophane/ or Insulin, Long-Acting/ or Insulin, Regular, Pork/ or Insulin/ or Insulin, Short-Acting/ or Insulin, Regular, Human/ or Hypoglycemic Agents/ or Thiazolidinediones/ or Sulfonylurea Compounds/ or (Metformin or (Insulin not ("insulin resistance" or "insulin receptor")) or Sulfonylurea* or Pioglit* or Thiazo*).ab.
2. epidemiologic studies/ or case-control studies/ or cohort studies/ or exp clinical trial/ or meta-analysis/ or exp epidemiology/ or ("case control" or "case - control" or "cohort" or "follow up" or "longitudinal" or hazard or risk or rate or odds).mp.
3. Diabetes Mellitus, Type 2/dt
4. Diabetes Mellitus, Type 2/
5. neoplasms/ or exp neoplasms by histologic type/ or exp neoplasms by site/ or exp neoplasms, multiple primary/
6. (DIABETES and ("type" and ("2" or II))).ab.
7. 5 or (cancer or tumor or malig* or neoplas*).ab.
8. 3 and 7
9. (4 or 6) and 1 and 7
10. (8 or 9) and human/
11. 10 and 2
12. 11 not (Cell line/ or cell proliferation/ or genetic association/ or genotypes/ or animal/ or cytokine.mp. or chemokine.mp.)
13. limit 12 to english
14. obesity/ or weight loss/
15. 13 not 14

OR

16. (("anti diabet*" or "anti-diabet*" or Metformin or Sulfonylurea* or Pioglit* or Rosiglit* Thiazo* or antidiab* or (insulin not ("insulin resistance" or "insulin receptor" or "insulin-like" or "insulin like"))) and (cancer or (tumor not "tumor necrosis factor") or malig* or neoplas* or carcin*)).ti.
17. 16 and 2
18. 17 not 15
19. 18 not (Cell line/ or cell proliferation/ or genetic association/ or genotypes/ or animal/ or cytokine.mp. or chemokine.mp.)
20. limit 19 to English

Box 2.1 Full MEDLINE search terms

During an initial title and abstract screen, reviews, meta-analyses and editorial pieces that looked at metformin and cancer were retained so that reference lists could be checked. Additionally, papers that appeared not to meet inclusion criteria (e.g. those that had primarily compared cancer incidence between diabetics and non-diabetics), were kept for full text screening in case the required measure of effect was reported as a secondary analysis. A full text screen was then applied to the remaining papers, and the reference lists of relevant reviews and meta-analyses searched. To test the reliability of the inclusion criteria an additional researcher screened a 10% random sample of the extracted studies. A Cohens kappa score was calculated to give a quantitative measure of rater reliability [44], with a value of 0.75 used as the threshold for "excellent agreement" [45].

2.2.3 Data extraction and bias assessment

The data extraction table was piloted on five studies, and subsequently refined to ensure systematic documentation of the relevant information. The key information extracted consisted of details on the data source, study population (including size and follow up length), study entry, study exit, exposure definition, exact outcome(s) studied, comparator group, considered covariates, analysis methods, details of any sensitivity analyses, key results and conclusions. Detailed criteria for assessment of bias were produced in order to consider risks of bias for each study. The 8 domains assessed for bias were 1) outcome definition; 2) exposure definition (including choice of comparator); 3) control selection (case control studies only); 4) consideration of HbA1c, BMI and other anti-diabetes drugs as time-dependent confounders affected by prior treatment; 5) adjustment for baseline (study entry) confounders (smoking, diabetes severity, age, gender); 6) immortal time (cohort studies only); 7) missing data; and 8) censoring methods (cohort studies only). For each bias domain, pre-defined criteria allowed categorisation into high, medium, low or unlikely risk of bias. These criteria are presented in appendix 4. Broadly, studies were considered unlikely risk of bias in a particular domain if the design and analysis methods were unlikely to induce a systematic difference between risk of cancer between metformin users and non-users. Low risk meant that there was small possibility of bias but the likely magnitude of the bias was not expected to have materially affected the overall study conclusions. Medium and high risk of bias meant that there was potential for moderate or substantial bias respectively in the study findings. Although the specific criteria for each bias domain may have left some room for subjectivity, they were developed in advance to make them as objective as possible.

Time-dependent confounders affected by prior treatment (e.g. HbA1c and BMI) were considered as a separate domain in addition to baseline confounding, to highlight the difference between baseline confounding that could be adjusted for in a standard analysis given appropriate data, and the subtler bias that may arise if valid methods to adjust for time-dependent confounders affected by prior treatment are not used. As an example, a study seeking to answer an "intention to treat" (ITT) question, where exposure was defined at baseline then assumed constant was considered at unlikely risk of time-dependent confounding by HbA1c even if it only adjusted for baseline HbA1c. This is because the exposure is no longer time-varying, so the issue of time-dependent confounding is eliminated. The same adjustment in a design where exposure was time updated after baseline was categorised as at risk of bias in the time-dependent confounding domain due to potential for confounding between post baseline changes in exposure and cancer. If studies omitted a particular confounder because they found it did not alter the estimate of metformin on cancer risk in a multivariable model, then they were not deemed to be at risk of bias due to its exclusion. However, the timing and accuracy of the confounder were still considered as sources of bias, since these aspects could have resulted in its incorrect omission. Bias from outcome and exposure definition encompassed both misclassification bias, biases induced by timing of measurement, and selection biases resulting from the definitions. Potential bias induced by using time-varying exposure without consideration for the time needed for exposure to plausibly cause cancer, could be considered as inappropriate censoring, or as inappropriate exposure definition; to avoid double counting this was considered a censoring bias.

Some studies provided multiple estimates based on dose response categories (13 studies), or differing comparators (5 studies). In this situation, the main estimate used for our analyses was that deemed to be most comparable to other studies. For multiple estimates from a dose response model, if an overall exposed vs non-exposed estimate was not presented (5 studies), a middle category best representing a moderate level of exposure was taken.

2.2.4 Meta-regression

As an exploratory analysis designed to investigate whether between study heterogeneity in the observed effect of metformin could be explained by bias and other study level factors; a random effects meta-regression was performed. Separate regressions were performed for the five most common outcomes – all cancer, colorectal/bowel cancer, lung cancer, breast cancer and pancreatic cancer. Studies that reported only stratum specific results (3 studies) were each

entered into a meta-analysis to generate a pooled estimate for that study, which was subsequently used in the meta-regression.

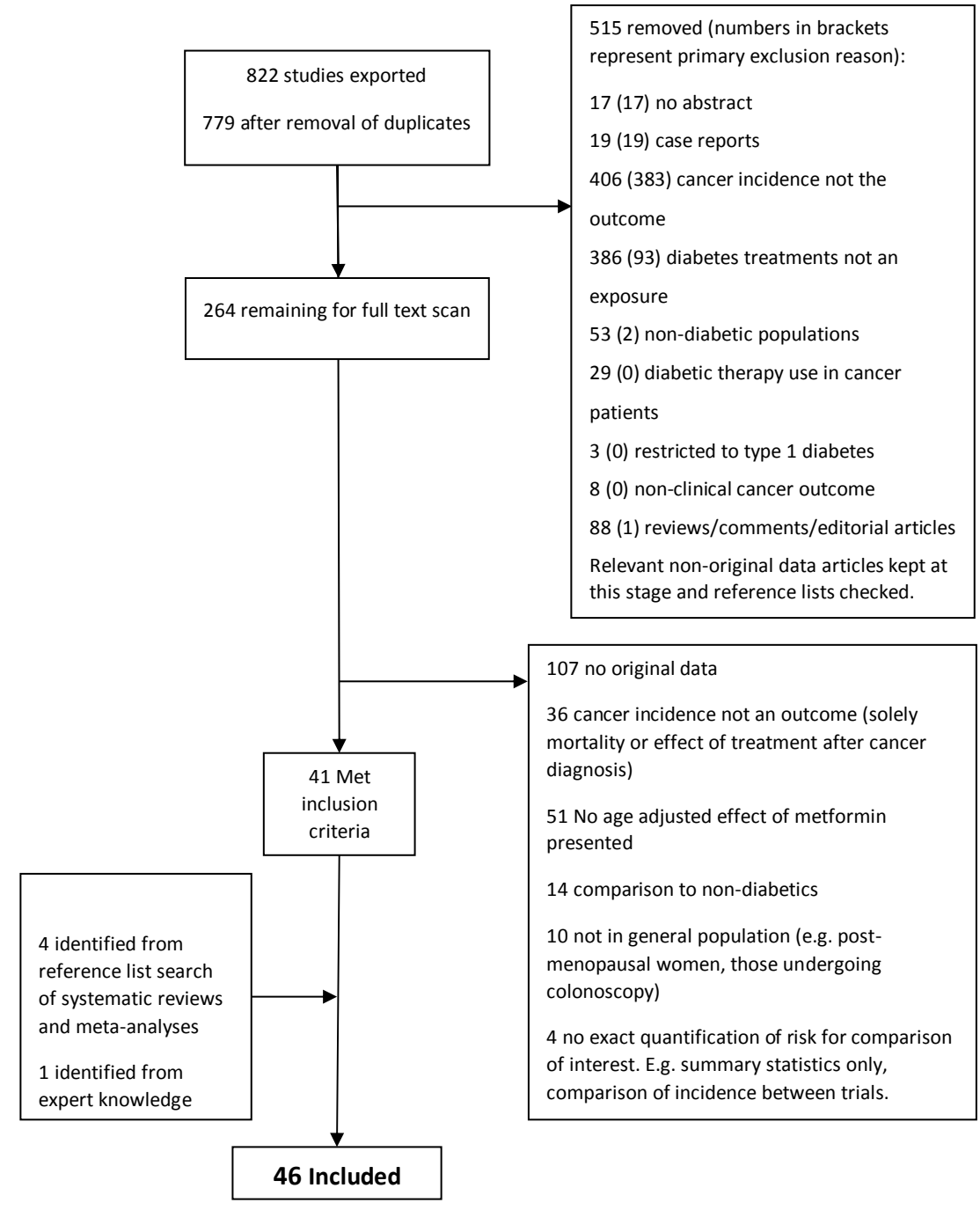
Study characteristics evaluated in the meta-regression were a subset of all available, based on a -priori assumptions about which might have the largest impact on study estimates. Characteristics included were comparator exposure (diet only, OADs, less metformin, and no metformin (diet and other OADs combined)), bias in exposure definition, bias in outcome definition, bias from baseline adjustments, bias from time-dependent confounders, immortal time bias, and whether the cohort were incident users of diabetic drugs. Zero was assigned to studies rated as unlikely or low in the bias assessment, and one to those rated medium or high. A binary classification was chosen to reduce sparsity. Backwards stepwise selection was used to identify which (if any) characteristics best explained study heterogeneity. A p value cut off of 0.4 was used due to small sample size and the large number of parameters in the full model.

2.3 RESULTS

2.3.1 Search and screening

The numbers of studies included/excluded at each stage of the process are presented in Figure 2.1. From an initial 822 references (779 after removal of duplicates), 46 studies were included in the final review. Full texts were available for all studies. The random sample of 76 studies independently screened by two researchers against the inclusion criteria resulted in a Cohens kappa of 0.79, and only a single initial disagreement over inclusion of a study; it was agreed on discussion that this study did meet the inclusion criteria. One article examined adverse event reports from two randomised controlled trials so was technically not observational, however it was included as it could be considered a retrospective cohort study with a trial based data source. It did not adjust for age but this exclusion criterion was waived since treatment was randomised.

Figure 2.1 Flow chart of screening process, detailing number of studies excluded at each stage and reason for exclusions



Flow chart as presented in [43] 10.1093/ije/dyw275

	Case Control N (%)	Cohort N (%)	Total N (%)
Number of studies	**22**	**24**	**46**
Data Source			
Clinical Trial	0 (0)	1 (4)	1 (2)
Diabetes Registry	2 (9)	4 (17)	6 (13)
Insurance database	2 (9)	9 (38)	11 (24)
CPRD (or GPRD)	8 (36)	6 (25)	14 (30)
Other primary/secondary care database	1 (5)	4 (17)	5 (11)
Recruited from Hospital/Clinic	9 (41)	0 (0)	9 (20)
Outcome definition*			
All cancer	5 (23)	16 (67)	21 (46)
Colorectal/Bowel	2 (9)	12 (50)	14 (3)
HCC/ICC	5 (23)	2 (8)	7 (15)
Ovarian/Endometrial	2 (9)	1 (4)	3 (7)
Bladder	0 (0)	3 (13)	3 (7)
Breast	3 (14)	10 (42)	13 (28)
Oesophagus	0 (0)	4 (17)	4 (9)
Kidney	0 (0)	2 (8)	2 (4)
Liver	0 (0)	5 (21)	5 (11)
Leukaemia	0 (0)	1 (4.2)	1 (2)
Lung	4 (18)	8 (33)	12 (26)
Melanoma	0 (0)	2 (8)	2 (4)
Pancreas	3 (14)	10 (42)	13 (28)
Prostate	3 (14)	8 (33)	11 (24)
Stomach	1 (5)	4 (17)	5 (11)
Definition of exposure to metformin for primary estimate			
Any Exposure	14 (64)	8 (33)	22 (48)
Any exposure but minimum time/number of prescriptions needed	1 (5)	2 (8)	3 (7)
Total Exposure (Number of prescriptions/time on metformin)	6 (27)	4 (17)	10 (22)
Monotherapy	1 (5)	8 (33)	9 (20)
Randomisation	0 (0)	1 (4)	1 (2)
Combination therapy with a sulfonylurea	0 (0)	1 (4)	1 (2)
Timing of Exposure measurement			
Current use (at time of cancer/matched date)	3 (14)	0 (0)	3 (7)
Time updated (current/ever/cumulative)	0 (0)	8 (33)	8 (17)
Fixed from start of follow up, with exposure occurring in a baseline period or follow up starting from first exposure (Intention to treat (ITT)).	0 (0)	8 (33)	8 (17)
Single summary measure of exposure over entire follow up.	19 (86)	8 (33)	27 (59)
Comparator group for primary estimate			
Less exposure (i.e. continuous exposure variable)	0 (0)	2 (8)	2 (4)
Diet Only	0 (0)	1 (4)	1 (2)
Rosiglitazone	0 (0)	1 (4)	1 (2)
Sulfonylurea	2 (9)	9 (38)	11 (24)
Any other OAD	3 (14)	4 (17)	7 (15)
No metformin (combination of diet and other OADs)	17 (77)	7 (29)	24 (52)
**** New users of Oral Antidiabetic Drugs (OADs)**			
Yes	3 (14)	7 (29)	10 (22)
No	17 (77)	12 (50)	29 (63)
Unsure	2 (9)	5 (21)	7 (15)

Table 2.1 Frequency table to summarise data source, outcome and exposure definitions for 46 studies.

*Studies may have multiple outcomes therefore column percentages will not sum to 1. **Based on whether clear description given in methods.
Table as presented in [43] 10.1093/ije/dyw275

2.3.2 Study characteristics

Table 2.1 summarises the data sources, outcomes, exposure definitions, timing of exposure measurements, and comparator exposures used. More detailed study level information is presented in appendix 5. Of the 46 studies, 22 were case control design [22, 46-66], and 24 were cohort studies [23-28, 67-84]. Thirty-seven (80%) of the studies used data from electronic health records; most commonly, the UK's Clinical Practise Research Datalink (CPRD) (13 studies) and the Taiwan National Health Insurance Claims Database (8 studies). As previously mentioned, one paper [72] used data from two randomised controlled trials. The remaining 8 (all case control) collected data from a specific cancer or diabetes clinic.

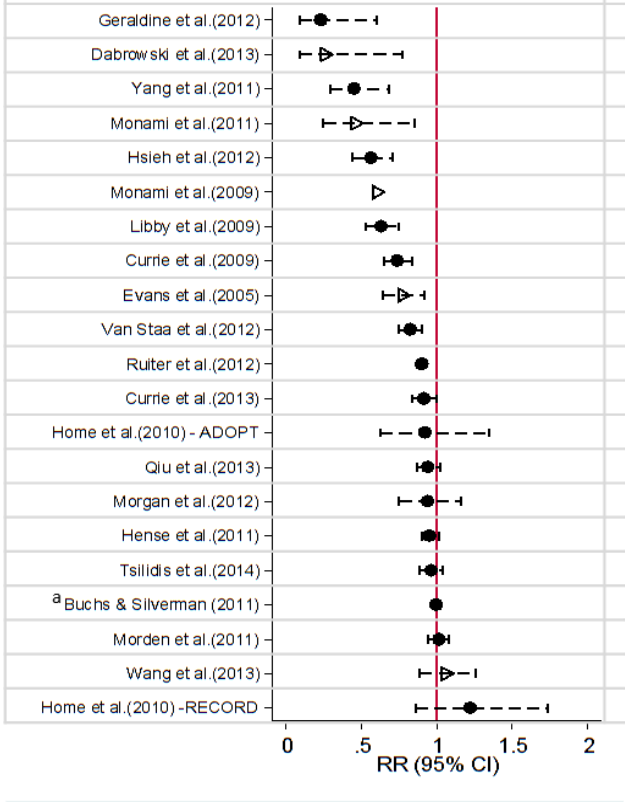
Twenty-two studies (46%) defined exposure to metformin as any exposure, without considering overall duration. Three further studies refined this definition by requiring a minimum time period or number of prescriptions before an individual was considered exposed. Nine studies (20%) looked at monotherapy with metformin and 10 studies (22%) used total exposure to enable dose-response analyses. The remaining 2 studies looked at metformin in combination with specific OADs, with a comparator group that allowed the estimation of the effect of just metformin. The most frequently used comparator group was no metformin, used in 24 studies (52%). Use of sulfonylureas (another popular first line oral agent) (11 studies (24%)) was also a common comparator.

There were 116 estimates presented for the effect of metformin on risk of cancer when considering separate estimates for different cancer sites. Twenty studies examined the outcome of all cancer excluding non-melanoma skin cancer (NMSC) (or similarly, a combination of a range of cancer types). One provided two estimates based on different data sources [72] making 21 estimates in total. Colorectal and/or bowel (14 studies) were the most common sites studied, followed by pancreas (13 studies), breast (13 studies), lung (12 studies) and prostate (11 studies). Other sites had less than 10 estimates each.

2.3.3 Effect of metformin on cancer risk

Figure 2.2 displays the study estimates and 95% confidence intervals (CIs) for relative risk (odds ratio (OR) or hazard ratio (HR)) of metformin use on incidence of all cancer. Estimates and 95% CIs for the four most commonly studied site specific cancers are presented in Figure 2.3.

Figure 2.2 Study specific estimated relative risk (odds Ratio or hazard Ratio) with 95% CI for metformin vs comparator on risk of all cancers, and corresponding assessment of risk of bias.

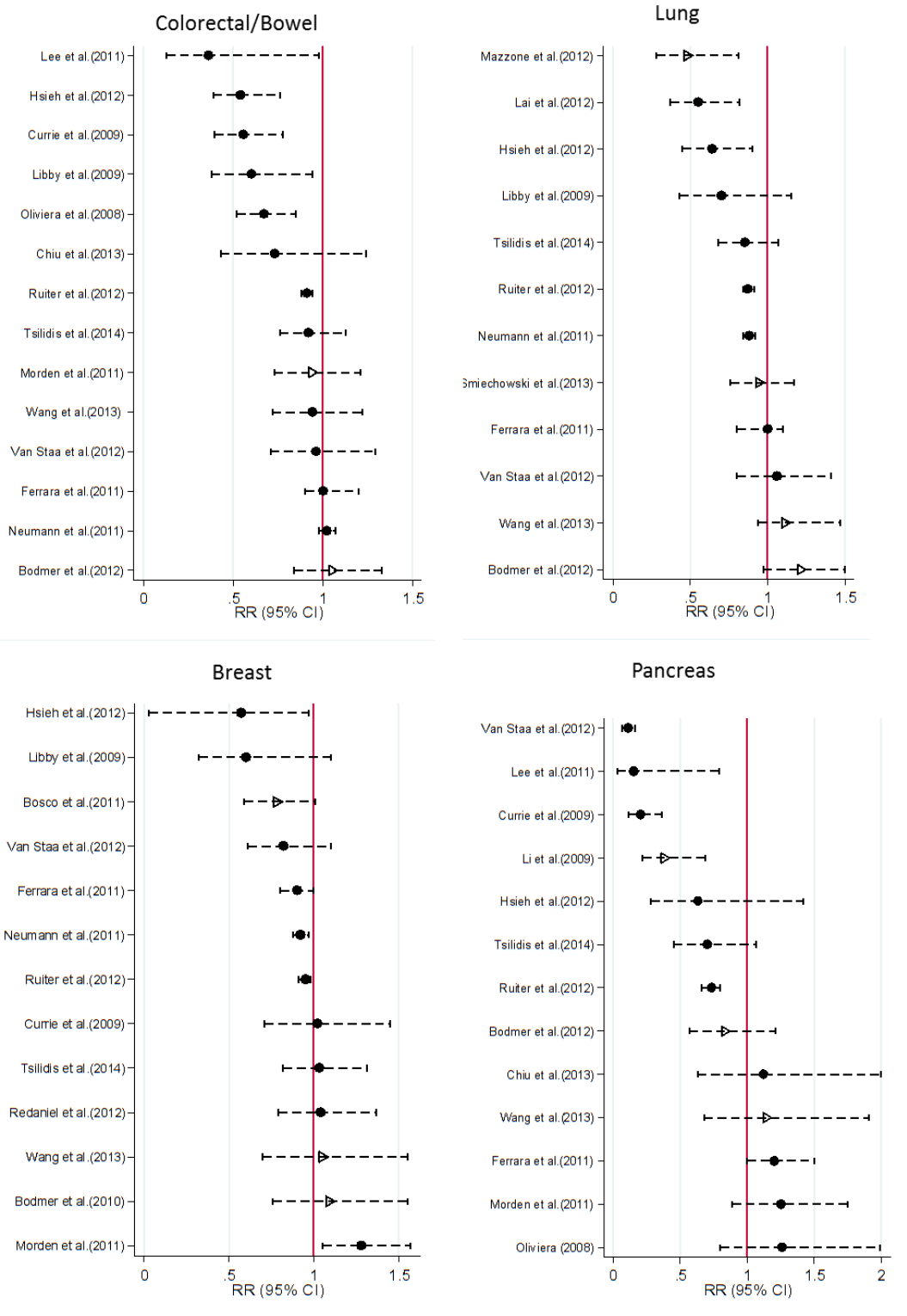


Outcome definition	Control Selection	Exposure definition	Treatment of HBA1c, BMI and Other OADs	Other Baseline Adjustments	Missing Data	Immortal Time	Censoring	Comparator exposure
✓		!!	✓	✓	?	!!	?	Diet
✓	!	!!	!!	!!	✓			No Metformin
✓		!!	✓	✓✓	?	!!	✓✓	No Metformin
✓	✓✓	!	!	✓	✓✓			No Metformin
✓		!!	!!	!!	✓✓	✓✓	?	Sulfonylurea
✓	!	!	!	✓✓	✓✓			No Metformin
✓		!!	!	✓✓	!!	✓✓	?	No Metformin
✓		!!	!	!	?	✓	!!	Sulfonylurea
✓	!	✓✓	!	✓	!!			No Metformin
✓✓		✓✓	✓	✓✓	✓	✓✓	✓	Less Metformin
✓✓		✓✓	!!	!	?	✓✓	!	Sulfonylurea
✓		!!	✓	✓	?	✓	!!	Sulfonylurea
✓✓		✓✓	✓✓	✓✓	✓✓	✓✓	!!	Rosiglitazone
✓✓		!	✓	✓	!	✓	✓✓	Sulfonylurea
✓		✓✓	✓✓	✓✓	?	✓	?	Sulfonylurea
✓		!	✓	✓	✓✓	✓✓	!	No Metformin
✓✓		✓✓	✓✓	✓✓	✓	✓✓	✓	Sulfonylurea
✓		!!	!	!!	✓✓	!!	!	Less Metformin
✓		!	!	✓✓	?	✓✓	?	No Metformin
✓	!	✓	!!	!!	✓✓			No Metformin
✓✓		✓✓	✓✓	✓✓	✓✓	✓✓	✓	Rosiglitazone

Risk of bias: ? Unknown !! High ! Medium ✓ Low ✓✓ Unlikely

[a] Represents the hazard ratio for cancer risk per one extra prescription of metformin. Figure as presented in [43] 10.1093/ije/dyw275

Figure 2.3 Study specific estimated relative risks (odds ratio or hazard ratio) with 95% CI for metformin vs comparator on risk of 4 most commonly studied site specific cancers



Case control studies are represented by hollow triangle, Cohort studies by filled circles. Figure as presented in [43] 10.1093/ije/dyw275

For all cancer, 18/21 estimates were below one, with 12/18 having upper confidence limits below or equal to one. The magnitude of the effect estimates ranged from just a 0.04% reduction in risk [67], to a 77% reduction in risk [70]. For site specific cancers, estimates were also highly variable across studies (Figure 2.3).

2.3.4 Bias evaluation

Study specific results of bias assessment for studies assessing all cancer as an outcome are displayed alongside risk estimates in Figure 2.2. Only three studies [28, 72, 83] scored low or unlikely for risk of bias in all categories. One further study, which looked at lung cancer only, scored unlikely or low in all categories except missing data [65], where it was rated unknown. Three of these studies saw no evidence of an effect of metformin. One study estimated a modest protective effect of long term use (>60 months) in comparison to short term use (0-6 months) with a hazard ratio (HR) of 0.82 0.75-0.90, but ultimately concluded that there was no evidence for a causal effect due to patterns of risk that were inconsistent with causality [83]. Of the 12 studies that estimated a statistically significant protective effect of metformin, eleven had at least medium risk of bias in at least two domains. Nine had medium or high risk of bias from exposure definition and seven had medium or high risk of bias for treatment of HbA1c, BMI and other OADs. The fully detailed bias assessments for each study are presented in appendix 6.

Time-dependent confounders affected by prior treatment

Table 2.2 (case control) and Table 2.3 (cohort) detail which adjustments were made for HbA1c, BMI and other anti-diabetes drugs, and the timing of the measurement within the follow up period for each study separately. Only four studies were considered as unlikely to be at risk of bias due to how HbA1c, BMI and other anti-diabetes treatments were accounted for in the analysis. These studies considered exposure to metformin as fixed from baseline ("intention to treat" (ITT) principle), and had confounders measured immediately prior to baseline.

Only 16/46 studies included HbA1c as a confounder in the final model. Six further studies reported considering it as a potential confounder, but did not include it in their final model due to lack of statistical significance [26, 67] for its association with cancer, or because it did not alter the estimate of effect of metformin in the multivariable model [28, 47, 49, 81]. All but one of these studies [28] were still considered at risk since it was questionable whether the HbA1c used was representative of HbA1c at the time of starting treatment. Twenty-six studies accounted for

BMI in their final model. In most case control studies, the measurement of HbA1c and BMI was prior to the date of cancer diagnosis (or matched date for the control) but it was rarely clear where this occurred in relation to the measurement of exposure. Therefore, the potential for these studies to have adjusted for factors on the causal pathway between metformin and cancer was high. For the cohort studies, most used BMI and HBA1c measurements at, or close to the time of cohort entry, which therefore either preceded or coincided with exposure classification. None of the studies reviewed used time updated values of either HbA1c or BMI, though some used averages across follow up. The appropriate adjustment for other anti-diabetes drugs is dependent upon the exposure and comparator group definitions. In six of 46 studies, adjustment for use of other diabetes drugs was not considered necessary [23, 24, 28, 72, 78, 81]; in the remaining studies, 22/40 accounted for it.

Study Name	HbA1c adjustment			BMI adjustment			Other Diabetic Medication adjustment		
	value prior to first exposure	value between exposure and index date [a]	value at index date [a]	value prior to first exposure	value between exposure and index date [a]	value at index date[a]	value prior to first exposure	value between exposure and index date [a]	value at index date [a]
Azoulay et al. (2011) [46]		✓			✓			✓	
Becker et al. (2013) [47]					✓			✓	
Bodmer et al. (2011) [51]		✓			✓			✓	
Bodmer et al. (2010) [52]		✓			✓			✓	
Bodmer et al. (2012) (Lung) [48]					✓			✓	
Bodmer et al. (2012) (Pancreatic) [50]					✓			✓	
Bodmer et al. (2012) (Colorectal) [49]					✓			✓	
Bosco et al. (2011) [53]									
Chaiteerakij et al.(2013) [54]									
Dabrowski et al.(2013) [56]								✓	
Donadon et al.(2010) [57]			✓			✓			
Li et al. (2009) [59]					✓			✓	
Evans et al. (2005) [22]					✓				
Hassan et al. (2012) [60]									
Margel et al. (2013) [61]								✓	
Mazzone et al. (2012) [62]		✓			✓				
Monami et al. (2009) [64]			✓			✓		✓	
Monami et al. (2011) [63]				✓				✓	
Smiechowski et al. (2013) [65]	✓[b]	✓		✓[b]	✓		✓[b]	✓	
Wang et al. (2013) [66]									
Chen et al. (2013) [55]								✓	
Donadon et al. (2010) - 2 [58]						✓			

Table 2.2 Adjustment method for key time-dependent confounders affected by prior treatment: Case control Studies.

Table as presented in [43] 10.1093/ije/dyw275 [a] Index date: time of cancer diagnosis/matched date for control. [b] Sensitivity analysis assessed difference between adjusting for covariates measured before exposure or between 1 year prior to exposure and index date.

Study name	HbA1c adjustment			BMI adjustment			Other Diabetic Medication adjustment		
	value at cohort entry (at time of or prior to first exposure)	time updated	average of values/ at any point after exposure	value at cohort entry (at time of or prior to first exposure)	time updated v	average of values/ at any point after exposure	value at cohort entry (at time of or prior to first exposure)	time updated	average of values/ at any point after exposure
Currie et al.(2009)[26]									
Currie et al.(2013)[25]	✓			✓					
Geraldine et al.(2012)[70]	✓			✓[a]					
Home et al.(2010)[72]									
Hsieh et al.(2012)[23]									
Lai et al.(2012) (HCC)[73]									
Lai et al.(2012) (LUNG)[74]									
Lee et al.(2011)[75]									✓
Libby et al.(2009[27]			✓			✓	✓[b]		
Qiu et al.(2013)[81]									
Redaniel et al.(2012)[82]			✓	✓					
Ruiter et al.(2012)[24]									
Tsilidis et al.(2014) [28]				✓					
Yang et al.(2011)[84]	✓			✓					✓
Buchs & Silverman (2011)[67]									✓
Oliviera et al.(2008)[80]									
Hense et al.(2011)[71]				✓			✓		
Chiu et al.(2013)[68]									
Ferrara et al.(2011)[69]	✓							✓	
Lehman et al.(2012)[76]			✓						
Morden et al.(2011)[77]	✓c			✓*					
Neumann et al.(2011)[79]								✓	
Van Staa et al.(2012)[83]				✓				✓	
Morgan et al.(2012)[78]	✓			✓					

Table 2.3 Adjustment method for key time-dependent confounders affected by prior treatment: Cohort Studies

Table as presented in [43] 10.1093/ije/dyw275

[a]weight used instead of BMI [b]measured within 3 months/1 year of cohort entry (either side of first exposure)

[c]diabetes complications used as proxy measures for severity Grey boxes indicate that adjustment not necessary. For HbA_{1c} and BMI, this was due to randomised treatment allocation. For use of other OADs, adjustment was not necessary if the study looked at incident users of diabetes medications and censored at change in medication.

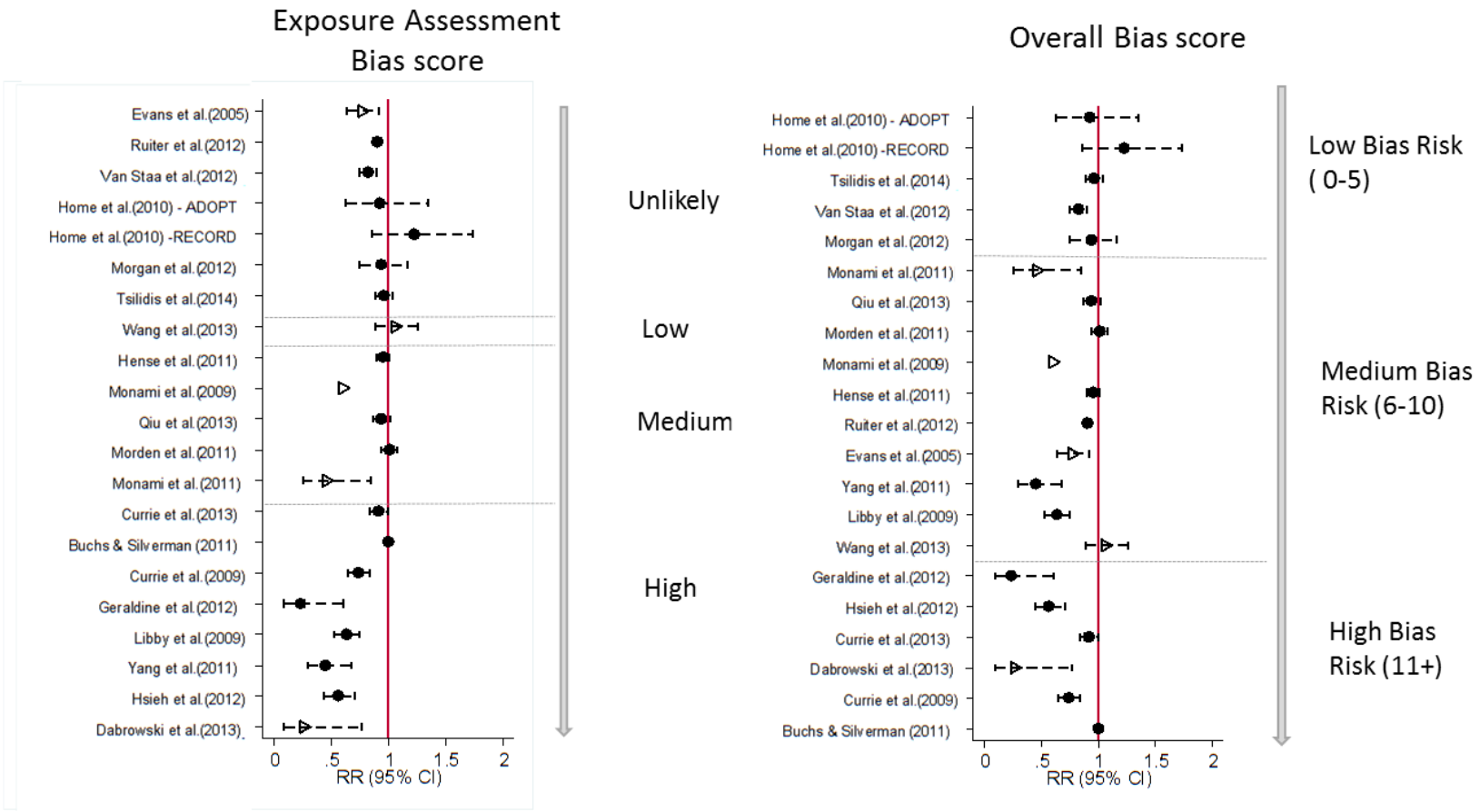
Other sources of bias

Exposure definition (n=28) and baseline adjustments (n=22) were the other most common reasons for medium or high risk of bias (see Figure 2.2 and appendix 6). The exposure definition was most likely to have introduced potential bias in case-control studies by having different time windows to measure exposure, meaning the overall chance of seeing individuals exposed to metformin was systematically different between the cases and controls. Potential bias was most often introduced into cohort studies because future information was used to inform exposure definition. 7/24 cohort studies were considered to have high risk of immortal time bias. 22 studies were considered at risk of bias from confounding due to incomplete or inappropriate baseline adjustment. This was because either the comparator used may have resulted in comparing patients at differing disease stages without adjustment for baseline disease severity, or because measures of severity used in the adjustment could be on the causal pathway between exposure and outcome, therefore not correctly accounting for differences in disease severity (and as such risk of cancer) that influenced choice of treatment at baseline. In addition, 36 studies were considered at risk of bias due to not considering a latency period for cancer (outcome definition). Since the effect of this bias is probably small in magnitude, this risk was considered low. This was supported by the five studies that considered different latency periods in sensitivity analyses, concluding that estimates did not differ substantially [24, 28, 46, 65, 82].

Many studies were considered to be at unknown risk of bias for censoring (12/24 cohort studies) and missing data (16 studies) due to a lack of information. Particularly for censoring, few cohort studies reported the numbers lost to follow up or for what reason. Four studies were rated medium or high for risk of bias from missing data, three of these because the missing indicator method was used, which will increase the risk of residual confounding [85]. With these three studies having > 20% missing data, the effect of residual confounding could be large.

By assigning values of 0, 1, 2 and 3 to unlikely, low, medium and high risk respectively, and summing over all domains, an overall bias score was calculated. When estimates for effect of metformin on all cancer were ordered by this score (Figure 2.4, right) it was observed that heterogeneity increased as risk of bias increased, and the strongest protective effects were observed in studies with the highest risk of bias overall. A similar figure was also created to look at bias from exposure definition only (Figure 2.4, left).

Figure 2.4 Estimates of relative risk of cancer with metformin use, ordered by risk of bias from exposure assessment only (left) and by overall risk of bias (right).



Overall bias score is sum of bias risk over all domains, with unlikely = 0, low = 1, medium = 2, high = 3. Case control studies are represented by hollow triangle, Cohort studies by filled circle. Figure as presented in [43] 10.1093/ije/dyw275

2.3.5 Meta-regression

Parameter estimates and model diagnostics for the final meta-regression models obtained are shown in Table 2.4. For the outcome of all cancer, after backwards stepwise selection, the only two study level predictors that remained in the model were comparator group and bias from exposure definition. The model estimated that using a comparator group of diet as opposed to no metformin made metformin appear more protective, whereas using other OADs or less metformin as a reference group made metformin appear less protective. Bias from exposure definition was estimated to make metformin appear more protective. However, the amount of residual heterogeneity between studies was still over 80%, and the estimated between study variance of this model was larger than the between study variance that would be estimated with no covariates in the model (as indicated by the negative adjusted R^2).

Both comparator group and exposure definition were also retained in the models for the site specific cancers, though in the models for colorectal/bowel, breast and lung cancer, using other OADs as the comparator was estimated to make metformin appear more protective. Bias from exposure definition was estimated to make metformin appear more protective in all site specific analyses. For example, for studies of lung cancer it was estimated to reduce the log risk ratio by 0.44, 95% CI (0.17, 0.72) p=0.007. For breast cancer, the strongest predictor was use of an incident user cohort, which made metformin look less protective. This predictor was also identified for studies of lung and pancreatic cancer, but the estimates had much less precision. Presence of both time-dependent and baseline confounding were also estimated to influence study heterogeneity for colorectal/bowel, breast and pancreatic cancer, with presence of these biases estimated to have broadly equal and opposite effects on the log risk ratio.

		All cancer		Colorectal/Bowel		Lung		Breast		Pancreatic	
		***Estimate 95% CI for effect on log risk ratio**	**P value**	***Estimate 95% CI for effect on log risk ratio**	**P value**	***Estimate 95% CI for effect on log risk ratio**	**P value**	***Estimate 95% CI for effect on log risk ratio**	**P value**	***Estimate 95% CI for effect on log risk ratio**	**P value**
Comparator Group	No metformin	0 (ref)		0 (ref)		0 (ref)		0 (ref)		0 (ref)	
	Diet only	-1.16 (-2.41 , 0.10)									
	Less Metformin	0.14 (-0.28 , 0.55)	0.217[d]	-0.04 (-0.43 , 0.34)	0.386[d]	0.05 (-0.34 , 0.44)	0.107[d]	-0.37 (-0.81 , 0.07)	0.625[d]	-1.66 (-3.30 , -0.01)	0.004[d]
	Other OAD	0.10 (-0.19 , 0.38)		-0.09 (-0.24 , 0.05)		-0.15 (-0.31 , 0.02)		-0.22 (-0.41 , -0.02)		0.36 (-1.10 , 1.83)	
Bias from Exposure Definition	Low Risk	0 (ref)		0 (ref)		0 (ref)				0 (ref)	
	High Risk	-0.16 (-0.43 , 0.10)	0.208	-0.40 (-0.58 , -0.21)	0.001	-0.44 (-0.72 , -0.17)	0.007			-0.84 (-1.74 , 0.06)	0.06
Bias from Outcome Definition	Low Risk					0 (ref)				0 (ref)	
	High Risk					-0.17 (-0.48 , 0.14)	0.234			-0.58 (-1.75 , 0.59)	0.238
Immortal Time bias	Low Risk									0 (ref)	
	High Risk									0.96 (0.03 , 1.90)	0.046
Bias from Time-dependent confounding	Low Risk			0 (ref)				0 (ref)		0 (ref)	
	High Risk			0.11 (-0.06 , 0.28)	0.171			0.22 (0.00 , 0.44)	0.047	0.92 (-0.13 , 1.96)	0.071
Bias from Baseline confounding	Low Risk			0 (ref)				0 (ref)		0 (ref)	
	High Risk			-0.12 (-0.18 , -0.06)	0.002			-0.22 (-0.46 , 0.02)	0.069	-1.05 (-2.01 , -0.10)	0.037
Incident users	Yes					0 (ref)		0 (ref)		0 (ref)	
	No					0.18 (-0.14 , 0.50)	0.218	-0.25 (-0.5 , -0.01)	0.041	1.28 (-0.20 , 2.76)	0.074
Constant		-0.15 (-0.43 , 0.13)	0.269	0.00 (-0.14 , 0.15)	0.954	0.01 (-0.16 , 0.18)	0.892	0.17 (-0.08 , 0.42)	0.155	-0.55 (-2.07 , 0.97)	0.371
[1]I squared		85.21%		0%		0%		3.5%		9.5%	
[2]Adjusted R^2		-20.32%		100.00%		100%		-190%		99.97	
[3]Tau^2		0.046		0		0		0.000305		0.000156	

Table 2.4 Parameter estimates from meta-regression models after backwards stepwise selection

*Estimate represents the expected change in the log risk ratio (either HR or OR) for the effect of metformin on cancer, for each study level predictor. [a] I squared is the estimate of residual variation due to study heterogeneity. [b] Adjusted R^2 is the estimated proportion of between study variance explained by the covariates in the meta-regression. This can be negative when the between study variation in the model is increased because of loss of degrees of freedom more than it is improved by the addition of the covariates. [c] Tau^2 is the estimate of the remaining between study variance. [d]Joint test of null hypothesis. Table as presented in [43] 10.1093/ije/dyw275

2.4 DISCUSSION

2.4.1 Key findings

This review has systematically identified and assessed the existing literature on the pharmacoepidemiological question of metformin use and cancer risk. The search identified a large number of studies from varying countries and journals, and the inclusion criteria were shown to have good reliability between raters. Although only one database was used in the search, by searching reference lists of other meta-analyses and systematic reviews the majority of studies are likely to have been identified, however the possibility that some relevant literature may have been omitted cannot be completely excluded.

The 46 studies examined in this review did not provide consistent evidence to support a protective effect of metformin on risk of cancer. Two of three studies with low or unlikely risk of bias for all categories had estimates consistent with no effect of metformin. The third study had an estimate consistent with a moderate protective effect with longer exposure to metformin, though the authors concluded this finding was unlikely to be causal [83]. In particular, this study included many analyses, and also reported that when comparing metformin exposure to other classes of oral antidiabetics, the risk of cancer did not differ between drugs. The authors also found that the incidence rates of cancer were higher in the first three months after therapy initiation which they suggested may be due to detection bias, which would also explain why longer exposure appears protective when compared to the first six months of therapy.

The estimates of effect reported across the 46 studies were highly variable for all outcomes studied. The bias evaluation performed was detailed and thorough, and every effort was made to agree in advance the criteria for risk of bias in each of the 8 domains examined. However, as in all studies of this kind, it was not possible to eliminate all subjectivity from this process.

Many studies were at high risk of bias from exposure definition, which, for reasons already outlined by Suissa & Azoulay [86], can have a large effect on estimates of risk. Within studies considered to be at low or unlikely risk of such bias, effect estimates were closer to the null, but there was still variation in point estimates albeit with some wide confidence intervals. It is possible that confounding by disease severity, and in particular confounding from time-dependent variables affected by prior treatment could partly explain the remaining heterogeneity in observed estimates.

2.4.2 Time-dependent confounding

Different approaches for dealing with a time-dependent confounder affected by prior treatment can have quite different consequences for interpretation. Figure 2.5A represents the total causal effect of metformin use on cancer risk that we wish to estimate in a simple example where we assume HbA1c is the only time-dependent confounder affected by prior treatment. Figure 2.5 B, C and D illustrate the causal pathways that are actually being estimated under the three approaches most commonly used in the studies examined in this review. In B, studies adjust for HbA1c but the measurement is taken anytime during follow up, which may result in "adjusting out" any effect of metformin that is mediated through HbA1c. Additionally, although not shown in the figure, if there are unmeasured confounders between HbA1c and the outcome, HbA1c acts as a "collider" [36]. This means that when HbA1c is adjusted for, an association between treatment and the unmeasured confounder is created, leading to further confounding of the association of interest. In C, because treatment may change after baseline, the single adjustment at time 0 may lead to residual confounding by post-baseline HbA1c. In D, the fixing of exposure from baseline removes the issue of time-dependent confounding and therefore allows the total effect of exposure on cancer to be estimated, but typically estimates an ITT effect only. This may not be appropriate given patients are unlikely to adhere to a single treatment throughout follow up.

One study adjusted for non-adherence [28] using a weighting method that produces an unbiased estimate if there are no unmeasured confounders of the association between non-adherence and outcome [87]. In this case, the validity of this assumption is questionable given that they only looked at time points of 1, 3 and 5 years post baseline, with an indicator for whether patients had switched treatment up to these times. By using the most recent covariates at each of these time points, the covariates used to predict switching may have already been affected by the switch. In general, the active comparator approach is also limited by only considering comparisons between drugs. When applied and analysed carefully, it will give an unbiased estimate of the effect of initiating metformin compared to initiating (as an example) sulfonylureas on development of cancer, however this is not necessarily equivalent to estimating causal pharmacological effect of metformin use on cancer incidence and may be inappropriate if the comparator in question may itself affect risk of cancer.

Most studies with low risk of other biases used the approach outlined in Figure 2.5D The lack of variation in how time-dependent confounders were adjusted for in these studies mean that it is

not possible with the current literature alone to assess whether there is a meaningful impact of time-dependent confounders affected by prior treatment on the estimated effect of metformin on cancer risk.

Figure 2.5 Directed Acyclic Graphs (DAGs) to represent estimated causal pathways for A) the desired total causal effect of treatment on cancer risk, and B)-D) the estimated effect under different methods of adjustment for time-dependent confounders affected by prior treatment.

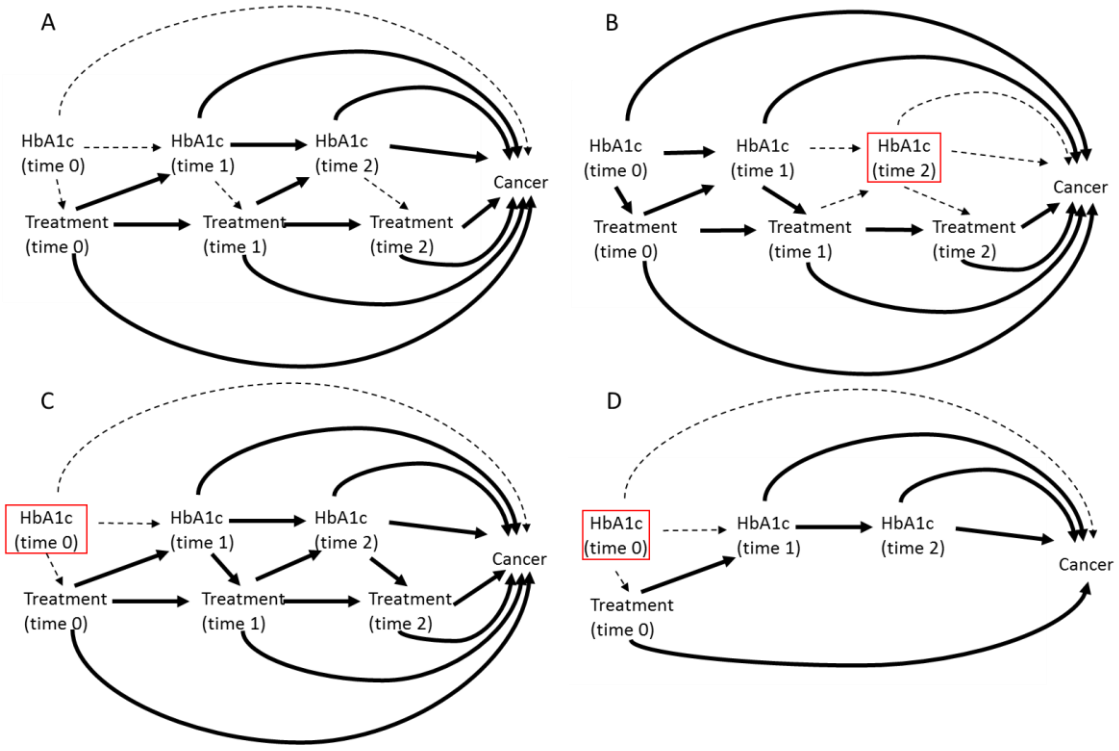


Figure as presented in [43] 10.1093/ije/dyw275
Box indicates adjustment.
Dotted line represents causal associations that are present but not included in the desired/estimated effect.
A Solid lines represent the pathways needed to estimate the total causal effect of time-varying treatment on cancer.
B HbA1c measured at a single time point during the measurement window (usually the most recent value). Exposure may be time updated or assumed fixed from cohort entry. Solid line represents the pathways included in the estimate of effect under this approach.
C HbA1c measured once at/before cohort entry, exposure modelled as time-varying. Solid line represents the pathways included in the estimate of effect under this approach.
D Exposure is assigned at cohort entry and assumed fixed (Intention to treat (ITT) principle), HbA1c measured once at/before cohort entry. Solid line represents the pathways included in the estimate of effect under this approach.

2.4.3 Meta-regression

The meta-regression aimed to establish whether any of the potential sources of bias could explain the heterogeneity in risk estimates. Although some of the results obtained (e.g. the estimated effects of risk of bias from exposure definition) are plausible, the overall reliability of the meta-regression is questionable. For all cancer there were 21 studies contributing to this analysis, and even after selecting only key study level predictors, there were nine parameters in the initial model. The analysis was likely underpowered and backward selection may not have produced reliable results. Additionally, many of these estimates lacked precision. For the site specific cancers, since the sample size for the meta-regressions were smaller, these issues may be enhanced further, and individual studies with extreme estimates are likely to have had a large influence. Furthermore, for some biases, two high risk studies could be rated as such for different reasons which would bias the estimate in opposite directions, resulting in the bias appearing to have no effect overall. In addition, the ability to examine only published studies may itself introduce a publication bias, which cannot be accounted for in a meta-regression. The lack of reliability is also compounded by the presence of negative values for the adjusted R^2, which indicate that the addition of parameters to the models has inflated the between study variance due to loss of degrees of freedom to a greater extent than the covariates improve the variance by explaining heterogeneity. Therefore, overall the results of this exploratory analysis should be interpreted extremely cautiously.

2.4.4 Conclusions

The existing literature provides inconsistent answers to the question of metformin use and cancer risk in type 2 diabetes. Variation in design of studies, and the potential for many kinds of bias make it difficult to fully explain the differences in risk estimates, particularly in terms of the potential impact of less easily detectable bias such as that from time-dependent confounders affected by prior treatment. It is likely that the largest protective effects that have been observed are a result of immortal time bias, and other issues relating to how metformin use was defined. Studies without such biases tend to have estimates closer to the null, and while an effect of metformin use on risk of subsequent cancer in patients with type 2 diabetes cannot be excluded, the previously reported large protective associations are unlikely to be causal.

In order to estimate the causal pharmacological effect of metformin on risk of cancer, the ideal would be to emulate a randomised controlled trial where patients are randomised to either metformin or diet only. This would involve comparison of patients initiating metformin with those controlling their disease by diet only, and correctly adjusting for disease severity at time of initiation while maintaining the effect of prior treatment on future disease severity measures. Using observational data, the only way to estimate such an effect would be with causal inference methods such as marginal structural models with inverse probability of treatment weighting or the g-computation formula, both of which will be introduced in chapter 4. This review and its update (see section 2.5) have shown that no studies to date appear to have used such causal methods to address the issue of time-dependent confounders affected by prior treatment in the metformin/cancer context. To be certain about the impact of such confounders, comparisons between adjustment methods need to be made within the same study population, using the same definitions of exposure, outcome, and confounders. A study utilizing causal methodology to deal with issues of time-dependent confounding, and how this compares to standard analysis methods would be a valuable and novel addition to the existing literature.

2.5 UPDATE ON STUDIES PUBLISHED SINCE THIS REVIEW

The same search as outlined in section 2.2.1 was repeated in November 2016, limiting to papers published since 2014 (to allow for studies that had been published but were not available on MEDLINE by May 2014). Forty-five studies were identified as potentially relevant new studies (excluding reviews and meta-analyses) from a title screen of around 350 additional articles. A more detailed screen of the abstracts and full text of these studies resulted in 24 new studies [88-111].

Key details of the 24 studies are presented in Table 2.3. None of the 24 studies used analysis methods that were substantially different than those covered in the original review. Some studies used propensity scores [93, 94, 107, 110] to deal with baseline confounding, either through inverse probability weighting [93, 110], or via adjustment for the score [94, 107]. The results of these studies were variable in terms of estimated effect of metformin, due to differing exposure groups, outcomes, and potential for immortal time bias in one study [94].

It was noticeable that more studies were explicit in stating that they used time-varying exposure to avoid immortal time, particularly when looking at cumulative exposure [90, 98-100, 103-109]. In contrast to the original review, there was also some variation in how time-dependent

confounders were adjusted for, with one study [90] time updating the adjustment for Hba1c as part of a secondary analysis. Time updating the adjustments for confounders will remove any effect of past treatment that acts via the confounder. It may also cause further confounding by inducing “collider bias” (as previously explained in 1.2). One study [99] adjusted for the mean value of HbA1c through follow up, however it not clear which pathways are (or are not) removed by such an adjustment.

Despite different methods of adjustment for time-dependent confounders, most of the newly identified studies using time updated exposure found little evidence of an association between metformin use and cancer risk, either when looking at ever/never medication, or cumulative exposure. Three studies focussed on risk of bladder cancer took different approaches to dealing with time-dependent confounding. Two studies time updated exposure, with one using the time updated adjustment [90] , and one with baseline adjustment as in Figure 2.5c [103]. A final study did not time update exposure (as in Figure 2.5d) [98]. The study using time updated adjustments [90] found a HR for metformin vs sulfonylureas for risk of bladder cancer of 0.92 (0.76-1.11). The study that time updated exposure with baseline confounder adjustment only [103] estimated the effect of metformin use vs non-use to be 0.6 (0.56-0.64). This study however did not adjust for HbA1c or BMI directly, instead using presence of comorbidities to measure diabetes severity. It is also not clear what the referent group were exposed to, and is likely a mixture of no drugs, and other first line alternatives. The third study that compared metformin to sulfonylureas in incident users only, with no time updating of exposure found an HR of 0.81 (0.60-1.09) [99]. These three estimates are somewhat consistent, though the point estimates do differ. Although the reference group in one of the studies is not perfectly comparable; the results from these three studies appear to support the view that the method of adjustment for time-dependent confounders may have impact on the estimated risk of metformin use on cancer.

Author	Year	Cancer type	Data source	Method	Comparison	OR/HR (95% CI)	Notes
Bosetti [88]	2015	HCC	Italian healthcare utilization databases	New user nested case control study	Metformin vs no metformin at cohort entry	0.67 (0.48 -0.95)	• Only a two year window to observe exposure and cancer, unlikely metformin can have an effect that quickly. • Not clear that diabetes severity has been balanced between those exposed and not exposed to metformin.
Cardel [89]	2014	Colorectal	Danish nationwide registries	Case control study	Metformin vs no metformin	0.83 (0.68 -1.00)	• Not clear that diabetes duration and therefore opportunity for exposure has been accounted for in this analysis. • Authors acknowledge that information on smoking, alcohol etc. is limited and that there may be residual confounding.
Goossens [90]	2015	Bladder	CPRD	New user cohort study / prevalent user sens analysis	Metformin vs sulfonylureas	1.03 (0.81,1.31) / 0.92 (0.76-1.11)	• Split follow up into 90 day periods and assigned exposure to each period as **current** if prescription in interval, **past** if not, or **none** if never prescribed. • Not clear what total person time/average follow up in the metformin only and sulfonylureas only categories is, so difficult to judge if long enough to observe cancer. • Authors also present analysis by time on metformin which gives an HR of 0.87 (0.46-1.63) for >=5 years use.
Hagberg [91]	2014	Liver	CPRD	Case control study	Metformin monotherapy vs no medication	0.74 (0.45-1.20)	• Adjusted for diabetes duration and matched on length of time in CPRD but cases not eligible to be controls for other cases before their index date. • All adjustments made prior to index date but this could be after metformin use so not clear that risk of cancer between exposed and unexposed fully adjusted for.
Kim [92]	2014	Gastric	Korean National Health insurance	Cohort study	Metformin vs no metformin	0.73 (0.53-1.01)	• Also report significant dose response effect by duration of metformin use. Info in table 1 suggests categorisation into use/no use and duration of use done using future information.
Ko [93]	2015	Endometrial	US healthcare claims	New user cohort study	Metformin vs sulfonylureas	1.09 (0.88-1.35)	• Censored at drug crossover. • Used inverse probability weighting of propensity score.
Kong [94]	2014	All site cancers	Hong Kong diabetes registry	Cohort study	Metformin vs no metformin	0.39 (0.25 , 0.61)	• Possible immortal time? • Authors present an analysis looking at statins and CVD, comparing different methods where immortal time is included/excluded. The model including immortal time gives closest result to the known effect of statins, justifying use of this model for their analyses.
Kowall [95]	2015	All site cancers	German Disease Analyzer database	New user cohort study	Metformin monotherapy vs sulfonylurea monotherapy	0.91 (0.73-1.14)	• Started follow up 1 year after first prescription. • As treated approach excludes monotherapy person time (I think) in patients that later intensify. • An ITT analysis where authors look at first medication and ignored changes was also conducted, gives an HR of 0.99 (0.85-1.13).
Kowall [96]	2015	All site cancers (plus some site specific))	Disease Analyzer database, Germany and UK	Multi database New user cohort study	Metformin monotherapy vs sulfonylurea monotherapy	1.02 (0.90-1.15)	• As above, ITT estimate 1.05 (0.99-1.12).

Table 2.5 Basic extraction information from the 24 additional studies identified in the updated search (Nov 2016)

Author	Year	Cancer type	Data source	Method	Comparison	OR/HR (95% CI)	• Notes
Lin [97]	2014	All site cancers	Taiwan National Health Insurance Database	Cohort study	Metformin monotherapy vs no medication	0.88 (0.77-1.01)	• Possible immortal time
Mamtani [98]	2014	Bladder	THIN	New user cohort study	Metformin vs sulfonylureas	0.81 (0.60-1.09)	• Patients censored at treatment crossover. • Time-varying cumulative exposure also examined. • Median follow about 2 years.
Onitilo [99]	2014	Breast, Prostate and Colon	Marshfield Clinic - healthcare system	New user cohort study	Metformin vs no metformin	0.65 (0.45 - 0.94)	• Time updated exposure - unexposed to exposed. • Adjustment for BMI and HbA1c is mean throughout follow up.
Sakoda [100]	2015	Lung	Kaiser Permanente diabetes registry	New user (of metformin) cohort study	Metformin vs no metformin	1.02 (0.85-1.22)	• Exposure time updated from never to ever for each medication class. • Mean of 9 years of follow up. • Also found no effect looking at time updated cumulative use. Some suggestion that metformin decreased risk in non-smokers HR 0.57 (0.33-0.99).
Sehdev [101]	2015	Colorectal	Marketscan	ase control study	Metformin vs no metformin	0.88 (0.77-1.00)	• Only looked at prescriptions for metformin in the 12 months prior to index date. Causal protective effect in such a short time unlikely. • Also found a non-significant protective effect of increasing metformin duration in days and by increasing dose.
Tsai [102]	2014	Lung	Taiwan National Health Insurance Database	New user cohort study	Years of metformin use (<1, 1-3, 3-5, >5) vs no use	<1 :0.69 (0.49-0.98) 1-3:0.55 (0.39-0.76) 3-5: 0.63 (0.47-0.86) >5: 0.76 (0.61-0.95)	• Possible immortal time .
Tseng [103-109]	2014	Thyroid	Taiwan National Health Insurance Database	Cohort study	Metformin vs no metformin at cohort entry	0.68 (0.60-0.78)	• Follow up in never users at baseline censored if they initiated metformin. If thyroid cancer can affect diabetes severity then many cancers that cannot be plausibly impacted by metformin may be missed.
		Bladder		New user (of metformin) cohort study	Metformin vs no metformin	0.60 (0.564-0.638)	• Time updated exposure, unexposed to exposed. • All adjustments made at baseline. (* using Propensity score adjustment) • Also looked at cumulative use but not clear that cumulative exposure was time updated. If updated from not exposed to the category of ever cumulative use then this will introduce immortal time? • History of cancer at time of diagnosis is much higher in those never treated with metformin for all studies. This could suggest post baseline confounding by previous cancers?
		Prostate*				0.47 (0.45-0.49)	
		Breast				0.63 (0.60-0.67)	
	2015	Endometrial				0.68 (0.61-0.74)	
	2015	Ovarian				0.66 (0.59-0.73)	
	2016	Kidney*				0.28 (0.25-0.31)	• As above, though inverse probability weighting of propensity score used. • Appears to be no adjustment for CKD/AKD, which is potentially important.

Table 2.5 continued: Basic extraction information from the 24 additional studies identified in the updated search (Nov 2016)

Author	Year	Cancer type	Data source	Method	Comparison	OR/HR (95% CI)	• Notes
Walker [110]	2015	Pancreatic	San Francisco Medical Clinics	Hospital based case control study	Metformin vs no metformin	1.01 (0.61-1.68)	• Also looked at months of metformin use ad found NS increased risk for all durations compared to no use.
Yen [111]	2015	Head and Neck	Taiwan National Health Insurance Database	New user cohort study	Metformin vs no metformin	0.66 (0.55-0.79)	• Not clear whether exposure status for controls based on entire follow up or just the first year. • Not matched on overall follow up so possible they are not exposed to metformin because they have shorter follow up…since rate of cancer higher later in time, this will produce a protective effect? Potentially issue with not matching on calendar time?

Table 2.5 continued: Basic extraction information from the 24 additional studies identified in the updated search (Nov 2016)

3 EFFECTS OF METFORMIN AND SULFONYLUREAS ON CARDIOVASCULAR EVENTS AND ALL-CAUSE MORTALITY

3.1 RATIONALE AND AIM

It is well established that type 2 diabetes confers an excess risk of cardiovascular (CV) complications [112]. Therefore, obtaining clear evidence as to whether particular treatment choices or strategies may affect this risk is of great importance. Outcomes such as MI, stroke, CV mortality, and combined endpoints such as major adverse cardiac events (MACE) and all-cause mortality have been frequently studied in both randomised clinical trials (RCTs) [31, 113-118] and various observational studies [32, 119] of oral anti-diabetes drugs. Some observational studies have suggested an increased risk to sulfonylurea users [33, 120, 121], particularly for all-cause mortality, compared to other treatment options and between different classes of sulfonylureas [122]. Since it is important to reduce any elevated risk of CV events as early as possible, comparisons of first line agents such as metformin and sulfonylureas are particularly pertinent. This section reviews the current literature relating to the effects of both metformin and sulfonylureas on the risks of CV events (limited to macro vascular events, in particular MI and stroke), CV mortality and all-cause mortality. Since first line therapy in newly diagnosed T2DM is the focus of subsequent chapters, the review is limited to any comparisons between metformin monotherapy, sulfonylurea monotherapy and no medication/diet only.

One of the main pathways by which type 2 diabetes medications may reduce the risk of CVD outcomes and all-cause mortality is by improving glycaemic control [123]. It is therefore important to know how well hyperglycaemia needs to be controlled to reduce this risk, and at what level of HbA1c a pharmacological therapy should be initiated over a lifestyle only intervention. UK guidelines currently have an HbA1c of 6.5-7% as the overall target and point of first line therapy initiation [13]. A recent article outlining the latest updates of evidence and policy [124] presents existing American target HbA1c values for managing CVD risk as <7% for most patients, <6.5% in patients with longer life expectancy, and <8% (or slightly higher) for older populations or where lifespan is limited by other comorbidities. However, the authors concluded that the epidemiological evidence to support these targets remains unclear. Therefore, as secondary aspect of this review, the existing evidence relating to target HbA1c control and levels for treatment initiation will also be summarised.

This review was less in depth than the previous review on metformin and risk of cancer. The reasons for this are firstly, that time did not allow another fully detailed systematic review, the scope of which would have been impractical with the given outcomes and quantity of existing research. Secondly, many review papers already exist for these questions and therefore there would have been considerable duplication of existing work. The aim was therefore limited to identifying and summarising the most important findings in this area.

3.2 SEARCH METHOD

A simple MEDLINE search using the following search terms was conducted to identify existing clinical trials and observational studies that have examined the effect of metformin or sulfonylurea use on cardiovascular events and glucose control.

(“metformin”OR”sulfonylureas”)AND(“type 2 diabetes”OR”type II diabetes”)AND(“myocardial infarction”OR”stroke”OR”all-cause mortality”OR”cardiovascular events”OR”cardiovascular risk”OR”glucose control”)AND(“epidemiology”OR”risk”)

This search produced approximately 600 results. A title screen identified 70 papers that would be potentially relevant. After reviewing abstracts, four of the most recent systematic review articles that presented results for the relevant comparisons were included. Two recent observational studies not included in the systematic reviews (identified by expert knowledge) were also obtained in full text. Finally, as a large long term clinical trial, the relevant publications from the UK Prospective Diabetes Study (UKPDS) were also obtained in full text. To identify whether any of the observational studies included in the systematic reviews used advanced statistical methodology to deal with issues of time-dependent confounding, the titles and study characteristics were reviewed via information provided in the review articles (including supplementary data), or by searching the studies and reviewing the abstracts. To identify papers relating to target glycaemic control or thresholds for treatment initiation; a recent umbrella review article was identified via a search of (“Glycaemic control”)AND(”cardiovascular risk”)AND(“type 2 diabetes”OR”type II diabetes”). This was used as a starting point to identify potentially relevant literature, with further studies identified via expert knowledge.

3.3 FINDINGS

A summary of the key studies included and the relevant reported results are organised by outcome and comparison in Table 3.1. The descriptive details of the different studies included are then presented by outcome.

3.3.1 Cardiovascular mortality

Three of the identified systematic reviews provided a summary of the existing evidence on use of metformin or sulfonylureas and the risk of CV mortality [30, 32, 120]. Two reviews included data from RCTs only, to examine the effect of sulfonylureas on CV mortality. The first review combined data from 18 trials [120], though only three trials compared to placebo/no therapy. The authors estimated an increased risk of CV mortality for sulfonylureas vs placebo, albeit with a very wide confidence interval, with a Mantel-Haenszel odds ratio (M-H OR) of 1.55 (0.17-13.64). The second review included 47 trials [30], but the summary measure for comparison to diet only/placebo was only based on three trials, with a summary estimate of 1.01 (0.68-1.51). Only one of these trials was included in the first review. Published in 2016, the third systematic review and meta-analysis reporting cardiovascular mortality as an outcome combined both RCTs and observational studies [32]. This review aimed to assess the evidence relating to a range of different medications, with a key comparison of interest being metformin monotherapy vs sulfonylurea monotherapy. For this comparison, the authors found that the range of risk ratios (RR) from the RCTs was 0.6 to 0.7 (two trials) and 0.6 to 0.9 for the observational studies (three studies), suggesting a reduced risk of metformin in comparison to sulfonylureas for CV mortality. For these particular outcomes, the authors did not present a summary estimate due to lack of studies. More studies (both observational and randomised) were identified but not included as they had either less than 1 year of follow up or were considered to be at risk of bias.

		Cardiovascular mortality		
Author (year)	**Type of Study**	**Met vs Diet**	**Sulf vs Diet**	**Met vs Sulf**
UKPDS 34 (1998)[31]	Clinical Trial	0.68 (0.53-0.87)		
UKPDS 33 (1998)*[114]	Clinical Trial		0.92 (0.68-1.23) 0.92 (0.69-1.24)	
Maruthur (2016) [32]	Review/Meta-analysis -Trials			Range 0.6-0.7 (2)
	Review/Meta-analysis - Observational			Range 0.6-0.9 (3)
Monami (2013) [120]	Review/Meta-analysis - Trials only		1.55 (0.17-13.64) (3)	
Varvaki Rados (2016) [30]	Review/Meta-analysis – Trials only		1.01 (0.68-1.51) (3)	
		Combined cardiovascular events		
		Met vs Diet	**Sulf vs Diet**	**Met vs Sulf**
Monami (2013) [120]	Review/Meta-analysis - Trials only		0.87 (0.91-1.07)	0.95 (0.34-2.70)
Morgan (2014) [33]	Cohort study			0.93 (0.88-1.14)
Lamanna (2011)*[34]**	Review/Meta-analysis - Trials only	0.79 (0.64-0.98) (8)		1.01 (0.82-1.23) (3)
Hippisley Cox (2016) [125]	Cohort study	0.68 (0.64-0.71)**	1.00 (0.95-1.05)**	
		MI		
		Met vs Diet	**Sulf vs Diet**	**Met vs Sulf**
UKPDS 34 (1998) [31]	Clinical Trial	0.61 (0.41-0.89)		
UKPDS 33 (1998)* [114]	Clinical Trial		0.87 (0.68-1.12) 0.78 (0.60-1.01)	
Pladevall (2016) [119]	Review/Meta-analysis -- Observational			0.80 (0.74-0.88) (4)
Monami (2013) [120]	Review/Meta-analysis - Trials only		0.82 (0.65-1.03) (3)	
		Stroke		
		Met vs Diet	**Sulf vs Diet**	**Met vs Sulf**
UKPDS 34 (1998) [31]	Clinical Trial	0.58 (0.29-1.18)		
UKPDS 33 (1998)* [114]	Clinical Trial		1.01 (0.65 - 1.58) 1.38 (0.52-2.08)	
		All cause mortality		
		Met vs Diet	**Sulf vs Diet**	**Met vs Sulf**
UKPDS 34 (1998) [31]	Clinical Trial	0.64 (0.45-0.91)		0.62 (0.40-0.98)*
UKPDS 33 (1998)* [114]	Clinical Trial		1.02 (0.82-1.27) 0.91 (0.73-1.15)	
Maruthur (2016) [32]	Review/Meta-analysis -Trials			Range 0.5-1.00 (2)
	Review/Meta-analysis - Observational			Range 0.5-0.8 (7)
Morgan (2014) [33]	Cohort study			0.77 (0.60-0.83)
Lamanna (2011)* [34]**	Review/Meta-analysis - Trials only	1.07 (0.56-2.06) (5)		
Hippisley Cox (2016) [125]	Cohort study	0.64 (0.63-0.66)**	1.24 (1.20-1.28)**	
Varvaki Rados (2016) [30]	Review/Meta-analysis – Trials only		0.97 (0.71-1.33) (3)	

Table 3.1 Summary table of included articles, type of study, and estimated effects of either metformin or sulfonylureas on multiple diabetes related outcomes.

For meta-analyses, number of studies contributing to summary estimate is given in brackets after the confidence interval.
Estimates are HRs/summary HRs unless otherwise stated.
*compared two difference classes of sulfonyulreas. Top estimate = chloropropamide vs diet, bottoms estimated = glibenclamide vs diet
**Although comparing to no therapy, the timing of no therapy in terms of diabetes duration was not clear. Therefore this could combine early diet treatment in relatively healthy patients, with later removal of treatment due to frailty.
*** 2 studies were in patients with impaired glucose tolerance rather than a formal diagnosis of T2DM.

Although not the largest, one of the RCTs with the longest follow up time (mean 10.6 years) examined in nearly all the reviews was the UK Prospective Diabetes Study (UKPDS) [31, 114]. Between 1977 and 1991, GPs in 23 centres across the UK referred patients with T2DM for potential inclusion. Patients between 25-65 years of age with FPG above 6.0 mmol/L were recruited. All patients received 3 months of conventional treatment with diet only. After this time, randomisation to two different sulfonylureas, insulin or continued conventional therapy occurred, with 619 patients randomised to chloropropamide and 615 to glibenclamide, 1156 to insulin and 1136 to remain on conventional treatment. In the first 15 centres, overweight patients could also be randomised to metformin. In total 1704 overweight patients were randomised, 411 to conventional therapy, 342 to metformin and 951 to one of the intensive treatments (either of the two sulfonylureas or insulin). This allowed for the effect of metformin vs diet only to be evaluated within overweight patients, and the effect of two classes of sulfonylureas to be examined vs diet only in an overall population of patients with T2DM. After 15 years of follow up, results suggested that compared to diet only, metformin monotherapy was associated with a reduced risk of diabetes-related death (similar to cardiovascular mortality) (HR 0.68 (0.53 – 0.87)). For sulfonylureas, there was no suggestion of a difference in risk compared to diet only, with HRs of 0.92 (0.68-1.23) and 0.92 (0.69-1.24) for treatment with chloropropamide and glibenclamide respectively.

3.3.2 Combined major cardiovascular events

More frequently, studies and reviews have examined occurrence of combined CV events. Lamanna (2011) [34] performed a meta-analysis of RCTs that looked at use of metformin in relation to risk of CV morbidity, MI, stroke, heart failure, all-cause mortality and CV mortality combined. With 12 studies included in the meta-analysis, they observed a summary OR of 0.94 (0.82-1.07).However, this included trials with differing comparator groups. When looking only at comparison to placebo or no therapy, eight studies contributed to a summary OR of 0.79 (0.64-0.98). For three studies where the effect of metformin was estimated by comparing metformin and sulfonylureas combined to sulfonylureas alone, no difference in risk of CV events was found (summary OR of 1.01 (0.82-1.23)). A Meta-regression of all 12 studies estimated that the effect of metformin on risk of CV events was estimated to be more protective in trials with longer duration and with a wider age range for inclusion.

The review by Monami [120] previously discussed for CV mortality, also examined the outcome of major adverse cardiac events (MACE). Two trials that compared sulfonylureas to placebo or

no therapy gave a summary OR of 0.87 (0.91-1.07), and a further two studies compared to metformin gave a summary OR of 0.95 (0.34-2.70).

An observational study by Morgan et al. [33] not included in any reviews also looked at risk of MACE (defined as acute MI or stroke) using data from the CPRD; by comparing new users of metformin with new users of sulfonylureas. The authors used propensity score methods in a sensitivity analysis to look at potential effect of insufficient control for confounding by indication. A small increased risk for sulfonylurea use was found the primary analysis (adjusted HR for metformin vs sulfonylureas 0.83 (0.76-91)) but in the propensity-matched cohort, the adjusted HR increased to 0.93 (0.88-1.14).

Another newly published study by Hippsley-Cox was a cohort study in UK primary care data [125], examining the effect of different diabetes medications on cardiovascular disease (CVD); defined as any of angina, MI, stroke or transient ischaemic attacks. They excluded prevalent users of insulin, gliptins, or glitazones, as well as anyone with a history of CVD or heart failure. Follow up time was split into treatment periods, allowing for periods of no therapy, monotherapy and combination therapy with different agents. They adjusted for a wide range of confounding variables at the beginning of each treatment period. Despite their main interest in comparing newer second line agents, results for exposure to a range of diabetes medications both alone and in combination were reported. The authors reported a decreased risk of cardiovascular disease when comparing metformin monotherapy to no treatment (HR 0.68, 0.65 -0.71). For sulfonylureas, the equivalent comparison gave an estimate of 1.00 (0.95-1.05). However, residual confounding may be an issue with this particular study. Periods of “no therapy” were not restricted to be in the early stages of diabetes, and so may include periods in which treatment was stopped due to high risk of death in the short term. In addition, the authors also adjusted for covariates at the beginning of each treatment period. This may have resulted in adjusting for covariates on the causal pathway between past exposure and outcome, or induced collider bias as previously explained in 1.2.

3.3.3 Myocardial infarction

One issue with combining cardiovascular events is that drugs may have different effects on individual events. In this case, it would be more appropriate to look at specific cardiovascular events in isolation. The UKPDS study reported events separately, finding that metformin was associated with a decreased risk of MI (HR 0.61 (0.41-0.89). For sulfonylureas vs conventional

therapy, the estimated HRs were 0.87 (0.68-1.12) for chloropropamide and 0.78 (0.60-1.01) for glibenclamide.

Three existing systematic reviews also reported results for MI alone [34, 119, 120], but one of these did not break down their overall summary HR by comparator group [34]. The meta-analysis by Monami [120] examined the effect of sulfonylurea use on MI, with their overall analysis containing 23 trials. When restricting this to comparisons to placebo or no therapy there were three studies, with a summary OR of 0.82 (0.65-1.03).

Pladevall (2016) [119] conducted a systematic review of observational studies that examined MI as an individual outcome, with four studies of metformin vs sulfonylureas included. These four studies combined gave a summary relative risk of 0.80 (0.74 -0.88). This review also assessed the potential for bias in each study, and it was observed that the study with the strongest protective effect of metformin vs sulfonylureas on risk of MI had the highest number of items at high or unclear risk of bias out of the four studies included.

3.3.4 Stroke

The UKPDS study was the only paper to present estimates for the effect of metformin or sulfonylureas on risk of stroke comparing to each other, or no therapy/placebo. For metformin, a protective effect was estimated but precision was low (HR 0.58 (0.29-1.18). Chloropropamide was not estimated to have an effect on risk of stroke compared to diet only, HR 1.01 (0.65-1.58). For glibenclamide, an increased risk was estimated but with very wide confidence intervals HR 1.38 (0.52-2.08).

3.3.5 All-cause mortality

For the effect of metformin or sulfonylureas on all-cause mortality, the only summary of evidence from observational studies comes from the review by Maruthur et al. [32] where seven observational studies comparing metformin use with sulfonylureas gave risk estimates from 0.5 to 0.8 for the relative risk of all-cause mortality. The corresponding range for RCTs was 0.5-1.02, based on two studies. Again, summary estimates were not presented for this outcome.

Two additional observational studies found consistent results. Morgan [33] estimated a decrease in relative risk for metformin vs sulfonylurea monotherapy of 33% (40% - 27%) in their

propensity matched analysis. In the Hippsley-Cox study, which examined the risks for metformin and sulfonylureas separately [125], use of metformin compared to no therapy was estimated to reduce the risk by 35% (HR 0.64 (0.63 – 0.66)). Sulfonylureas were estimated to increase risk of death vs no therapy (HR 1.24 (1.20 – 1.28)). However, as previously discussed in 3.3.2, this study had two notable limitations and the observed results should be cautiously interpreted.

Two further systematic reviews contributed to the evidence for metformin vs diet, and sulfonylureas vs diet only for risk of all-cause mortality. For metformin, the review of clinical trial data by Lamanna [34], estimated the summary OR for metformin vs placebo on all-cause mortality to be 1.07 (0.56-2.06), which is less supportive of a protective effect. The meta-analysis by Varvaki Rados [30] estimated an effect of sulfonylurea use vs diet only on all-cause mortality of 0.97 (0.71 - 1.33) based on three studies. Although this does not exclude a harmful effect, it is not consistent with the result from the Hippsley – Cox study.

The UKPDS study estimated a protective effect of metformin on risk of death vs diet only, with an HR of 0.64 (0.45-0.91). The equivalent results for the comparisons between chloropropamide and glibenclamide were 1.02 (0.82-1.27) and 0.91 (0.73-1.15) respectively. In a sub study of the overweight patients; patients randomised to metformin were subsequently randomised to combine with sulfonylureas or not. There was less follow up for this group as they were randomised six years after their initial randomisation to metformin, but comparing metformin only to metformin plus a sulfonylurea resulted in an HR of 1.60 (1.02-2.52).

3.3.6 HbA1c control and thresholds for treatment initiation

A recent umbrella review by Rodriguez-Gutierrez et al. [126] aimed to compare the latest guidelines on glycaemic control, with the latest evidence on “tight glycaemic control”. They examined five key RCTs [31, 114, 116, 117, 127], their extension studies (if conducted), 11 meta-analyses, 16 guidelines and 328 statements relating to glycaemic control in patients with T2DM.

The authors found that there was no evidence that tight control reduced risk of all-cause mortality or stroke, but they did observe a consistent reduction in non-fatal MI across all trials and reviews. In a meta-analysis of the trial extension studies, they estimated an overall RR of 0.85 (0.74-0.96) for the effect of “intensive control” (defined as discussed below) on non-fatal MI. The equivalent estimates for all-cause mortality and stroke were 1.02 (0.91 – 1.14) and 0.99 (0.89-1.08) respectively. They also found that the published statements and guidelines favouring tight glycaemic control for reduction of macro vascular endpoints had decreased with time. This

likely reflects the uncertainty that remains around this issue, making it difficult to inform policy in terms of target HbA1c.

To gain further insight into the evidence examined in the large scale umbrella review, two relatively recent and frequently cited meta-analyses of intensive glucose control and CVD risk included in the review were also examined. Ray (2009) [128] conduced a meta-analysis of 5 large RCTs, with the aim of understanding whether intensive glucose control was beneficial in terms of all-cause mortality and macro vascular events. Hemmingsen (2011) [129] conducted a similar review with 14 trials included. Both reviews concluded that the existing evidence suggest no decreased risk of all-cause mortality with intensive glucose control. For MI, both studies found similar overall ORs favouring intensive glucose control (summary ORs 0.83 (075-0.93) [128] and 0.85 (0.76-0.95) [129].

The five key trials discussed by Rodriguez-Gutierrez et al were included in both of these reviews. One issue with combining the evidence from these trials in meta-analyses is that they had differing definitions of “intensive” control. Three of the five studies examined by Ray randomised patients to treatment strategies that involved a target HbA1c. “Intensive” target definitions included 6%, 6mmol/L (5.4%) and 6.5%. These were then compared to “standard” treatment targets, which varied by definition (e.g. 7-7.9%, “as per local guideline”, or 15mmol/L (11%)). Therefore, despite the existing evidence comparing “intensive” vs “standard” control, the evidence to support a particular target HbA1c is still lacking. In addition, the UKPDS study was the only study in newly diagnosed T2DM. All other studies were conducted in patients with a mean of 8-10 years since time of diagnosis, meaning that the validity of the combined results in terms of early stage diabetes and the effect of early HbA1c control is questionable.

In order to identify an optimum target to reduce risk of all-cause mortality, Arnold and Wang [130] used data from six observational studies, identified via a systematic search, that examined the relationship between HbA1c and all-cause mortality. They used meta-regression techniques to estimate an overall association, finding a J shaped curve with minimum risk between 6% and 7.5%. This shape was only evident in an analysis restricted to studies that examined more than five HbA1c categories. In another analysis of 14 studies with more than two levels of HbA1c included, the authors did not observe this non-linear association, finding instead suggestion of increasing risk with increasing HbA1c. This study therefore provides some support for the current guidelines of targets between 6.5 and 7% but may also suggest that a target up to 7.5% could be considered without compromising overall mortality.

An alternative way to approach the issue of optimal HbA1c control may be to ask a slightly different but related question. Particularly relevant in newly diagnosed diabetes, where patients may present with a wide range of elevated HbA1c levels is the question of when to start pharmacological treatment over a lifestyle intervention. Looking at treatment initiation thresholds has the added benefit of being able to ask other useful questions, such as does it matter if we start treating at 10% or 7%, in terms of subsequently achieving "target" Hba1c? Alternatively, does starting first line therapy at a higher HbA1c affect long-term risk of cardiovascular outcomes? Comparing treatment thresholds in observational data is challenging, because the value of Hba1c at which a particular patient initiates treatment will depend on many factors that a) vary with time, b) are affected by past decisions to remain untreated, and c) affect future risk of CVD and mortality. Therefore, issues of time-dependent confounding arise.

One option to avoid time-dependent confounding would be to compare HbA1c levels at the time of treatment initiation for risk of cardiovascular outcomes. However, this approach does not fully compare treatment strategies relating to a threshold HbA1c, since they ignore person time in which patients are maintaining an HbA1c below the threshold but have not initiated pharmacological treatment. No existing literature could be found directly comparing HbA1c thresholds for first line treatment initiation on risk of long-term outcomes or on future HbA1c control. One study was identified that compared HbA1c levels at which treatment was intensified from metformin to a combination therapy, in terms of risk of MI, renal function, and albuminuria [131]. The authors used dynamic marginal structural models to compare HbA1c thresholds while adjusting for time-dependent confounding and informative censoring, using data from an American insurance claims database. They found that there was no increased or decreased risk of MI with lower HbA1c thresholds, but found some protective effect of lower HbA1c thresholds on reducing onset or progression of albuminuria.

3.4 DISCUSSION

The results summarised above appear to indicate that metformin has an overall protective effect on cardiovascular outcomes, both in comparison to sulfonylurea monotherapy, and to diet only. The evidence for the effect on all-cause mortality appears less conclusive. However, for comparisons to diet only, these findings (for all outcomes) are largely based on a single clinical trial (UKPDS), a single meta-analysis that only included 8 trials (5 for all-cause mortality) and included the UKPDS, and a cohort study that had an unclear risk of bias.

Four articles that compared metformin to sulfonylurea monotherapy suggested a protective effect of metformin for at least one outcome. This was most clear in observational studies or reviews of observational studies where, across all outcomes, 4/6 estimates had upper confidence limits below one. In contrast, the equivalent number for studies including clinical trial data was 2/5 (with a third study having an upper limit of 1 exactly). However, all that can really be established from these studies is there may be a difference in effect between the two drugs. The cause of the difference cannot be determined.

It was apparent that although the reviews of RCTs had a large number of trials included, comparisons of metformin and sulfonylureas to placebo or diet therapy were relatively limited. In particular, many RCTs had small sample size or short follow up. In order to establish the reason for observed differences in effect when metformin is compared with sulfonylureas, studies in large cohorts, with longer follow up comparing both metformin monotherapy and sulfonylurea monotherapy with diet alone are needed. Such comparisons in observational data, if not done with careful consideration and appropriate methods to deal with time-dependent confounding, have the potential to be biased. None of the existing studies identified, through the reviews or otherwise, appear have taken account of time-dependent confounders affected by prior treatment using appropriate causal methodology. A separate review by Patorno published in 2014 [132] identified 81 observational studies examining association between glucose lowering medications and cardiovascular outcomes, and aimed to describe the most common methodological limitations. Key issues identified included immortal time, adjusting for intermediate factors ("over adjustment bias"), censoring at treatment changes with no latency period, and confounding by disease severity. These conclusions are similar to those established in the review of studies of metformin and cancer in chapter 2, and give further support for the overall need for careful design and analysis when studying the effects of time-varying diabetes treatments.

One RCT with mean follow up on 10.6 years exists (the UKPDS study) that compared first line treatments to diet only. This study began in 1977, and there have been many advances in overall management of cardiovascular disease since then, which, as discussed be Ferrannini (2105) [133] may affect generalisability of such a study to the present day. Having said this, as long as one is mindful of these limitations, the results from the UKPDS study may serve as good comparators for results from an observational study of metformin and sulfonylureas vs diet alone. Specifically, this study found (i) evidence that metformin was protective for all of MI, stroke, and all-cause mortality, estimating a 30-40% decreased risk vs diet only for all outcomes;

(ii) no clear evidence that sulfonylureas were protective for any of MI, stroke or all-cause mortality, but also no strong suggestion that they were harmful.

In terms of the literature relating to optimal HbA1c control, this review suggested that intensive vs standard HbA1c control has been studied extensively in T2DM. More intensive control appears to lower the risk of all-cause mortality according to clinical trial data, however the actual optimal target for reducing long-term risk remains unclear. In addition, no literature was found directly comparing HbA1c thresholds for initiation of first line therapy. In chapter 9, such a comparison using observational data will be presented.

This was not a comprehensive systematic review, and by relying primarily on a few recent systematic reviews and meta-analyses, it is unlikely that all relevant literature has been identified. However, the main aim was to gain a general overview of the existing literature and weight of evidence for the effects of metformin, sulfonylureas and HbA1c control on cardiovascular outcomes to inform analyses presented later in the thesis. It was noticeable that, as eluded to above, many more trials and observational studies (and reviews of them) exist that have different comparator groups or look at the effect of treatments at later stages of disease [29, 121, 134]. However, as with the comparison between metformin and sulfonylurea monotherapy, the estimate of comparative effectiveness cannot be interpreted in terms of the effect of a specific drug, unless it is known that the comparator agent has no effect on the outcome of interest.

4 OUTLINE OF STATISTICAL METHODOLOGY

4.1 BACKGROUND

Causal models aim to provide estimates for the effect of exposure on outcome that can be given a causal interpretation. A causal estimate is the target of almost every study with a question such as "what is the effect of a on b?" whether it is a randomised trial or an observational study. The validity of the study result as a causal estimate invariably relies on certain assumptions, which vary with the context and methodology used.

This section will outline the formal definitions of a causal effect, and define other key terms in causal inference. It will then introduce the possible approaches for obtaining causal effect estimates in the presence of time-dependent confounders affected by prior treatment (see 1.2), and lay out the assumptions under which a causal effect may be established. The methodology of the specific modelling approach to be used in this thesis, namely inverse probability of treatment weighting of marginal structural models, will then be explained in further detail. An overview of the practical implementation will also be given. Further details relating to the practical implementation will be given in the methods section of the relevant analyses (chapters 7,8 and 9).

4.2 BASIC DEFINITIONS

4.2.1 Definition of a causal effect

Throughout, capital letters represent random variables and lower case letters represent observed values. Define A as a random variable representing the exposure of interest (usually treatment, which will be assumed to be the exposure from here onwards), where A can take the value $a = 0$ or $a = 1$, i.e. treated or untreated. If Y represents a continuous outcome of interest, then Y_a is the outcome if the treatment variable takes the value a.

It then follows that there are two **potential outcomes** for a subject, also known as counterfactuals, Y_0 and Y_1 that could be observed dependent on treatment. In real life, a subject is either treated or untreated, and can only experience one of these outcomes. These concepts,

and the following definitions were introduced and developed by Neyman et al [135] and Rubin [136] respectively.

Using the idea of potential outcomes, the individual causal effect of treatment *A* on outcome *Y* can be defined as

$$Y_1 - Y_0$$

In general, it is not possible to identify an individual causal effect. Instead, the average causal effect, which contrasts the expected potential outcomes for multiple observations, can be defined as

$$E[Y_1] - E[Y_0]$$

where $E[Y_a]$ is the expected value of the outcome Y where everyone's treatment status is equal to a. If Y is binary, the average causal effect may be re-defined as a risk difference, a risk ratio or an odds ratio.

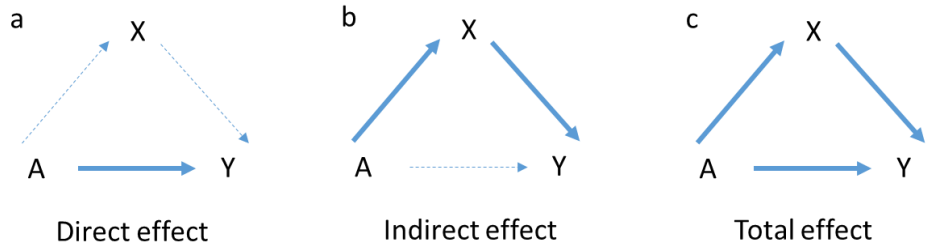
In a longitudinal setting, A can be represented as a vector $\underline{A}$ of longitudinal treatment choices, with $\underline{a}$ being the vector of observed values $a_1, a_2, a_3 \ldots a_T$ for time $t = 1,2,3, \ldots T$. The counterfactual outcome for a specific longitudinal treatment strategy is then represented by $Y_{\underline{a}}$.

4.2.2 Direct and indirect effects

A variable X on the causal pathway between A and Y, is a variable such that the effect of A on Y acts through X. Such variables may also be referred to as "mediators" [137].

When there are multiple pathways through which A can act on *Y*, the average causal effect can be dissected into **direct** and **indirect** causal effects, which sum to the **total effect**. The causal diagrams in Figure 4.1 below provide simple examples of each of these in the presence of a single mediator.

Figure 4.1 Simple depiction of direct, indirect and total effect of treatment A on outcome Y in the presence of a single mediator X.



Thick lines indicate the pathways included in the effect estimation

4.3 AN OVERVIEW OF METHODS FOR DEALING WITH TIME-DEPENDENT CONFOUNDING

As previously explained in 1.2; a variable that is both a mediator and a confounder when modelling time-varying treatment in longitudinal studies is known as a "time-dependent confounder affected by prior treatment", and standard statistical methods cannot estimate the total effect of time-varying treatment on outcome without bias in the presence of such confounders. Throughout this thesis, the term "causal effect" is taken to mean the total effect.

Three approaches exist that allow, in theory, the estimation of the total effect of time-varying treatment in the presence of time-dependent confounders affected by prior treatment. These are g-computation [38], inverse probability of treatment weighting of marginal structural models [37] and g-estimation of structural nested models [39].

4.3.1 G-computation

G-computation was introduced by Robins [38] in 1989, and is, as explained by Daniel et al [138], a generalisation of standardisation to the longitudinal setting. For a time fixed confounder, one could standardise over a confounder L by multiplying the expected outcome Y conditional on L by the probability of L, for each possible value of L, and summing the joint probabilities.

Longitudinally, the confounder L will have an observed vector of values $\underline{l}$, taking the values $l_1, l_2, l_3 .. l_T$ at time $t = 1,2,3, \dots T \dots$

G-computation allows the estimation of the expected potential outcome $E[Y_{\underline{a}}]$ of a particular treatment strategy $\underline{a}$, by expressing the expectation as a function of the conditional probability

of Y given treatment history $\underline{a}$ and covariate history $\underline{l}$; and the probability of $\underline{L}$ conditional on treatment and covariate history.

More formally, defining $A(t)$ to be the treatment at time $t = 1 \ldots . T$, and $\underline{A}(t^-)$ to be the vector of treatment up to time $t-1$, and defining $L(t)$ and $\underline{L}(t^-)$ analogously for a set of time-dependent confounders L (where we now allow L to mean a **set** of variables without changes to notation); the specification of the g-computation formula (or "g-formula") is, as in [138]:

$$E(Y_{\underline{a}}) = \sum_{\underline{l} \in \underline{L}} \{ E(Y | \underline{A} = \underline{a}, \underline{L} = \underline{l}) \prod_{t=1}^{T} P(L(t) = l(t) | \underline{A}(t^-) = \underline{a}(t^-), \underline{L}(t^-) = \underline{l}(t^-)) \} \quad (1)$$

This can be extended to continuous forms of L by replacing the sum with an integral (as shown in [138]). Due to the complexity for long follow up and large numbers of covariates, the practical implementation of estimating and contrasting potential outcomes under different treatment patterns uses Monte Carlo simulation to estimate the expected outcomes. Macros are available in both SAS and Stata for its implementation [139, 140].

Some examples of the application of the g-formula in existing literature include estimating the effect of hypothetical lifestyle interventions such as smoking cessation and increased exercise on risk of coronary heart disease (CHD) [139] , the effect of a new drug on mortality in bone marrow transplants [141] and the effect of asbestos exposure on lung cancer mortality [142].

4.3.2 Inverse-probability of treatment weighting of marginal structural models

Marginal structural models (MSMs) directly model counterfactual outcomes. They are so called because they describe the average causal effect of treatment on the marginal distribution of the potential outcomes, and because models for potential/counterfactual outcomes are referred to as "structural" in social and economic literature [143].

The most commonly applied approach to estimate the parameters of an MSM in the presence of time-dependent confounding, is that of inverse probability of treatment weighting (IPTW)) of MSMs [37]. Also developed by Robins [144], Daniel et al. describes the basic intuition behind IPTW by noting that it aims to

> "Re-weight the subjects in the analysis to mimic a situation in which the assignment to treatment is at random" ([138] page 1598).

The process of applying MSMs with IPTW is a two stage process. Firstly, each individual's probability of having their own treatment history is calculated; and used to calculate the IPTW. In the second stage, the treatment-outcome association is estimated in a regression model that is weighted using the IPTW.

The resulting estimate can then be interpreted as the estimate that would be obtained in a pseudo population where treatment allocation is independent of the time-dependent confounders.

Defining $A(t)$, $\underline{A}(t^-)$, $L(t)$, and $\underline{L}(t^-)$ as before, namely, as the value of treatment at time t, the full treatment history to time t-1, and the analogous variables for a set of time-dependent covariates L, the simplest IPTW is the "unstabilised" weight, defined as in [145], as:

$$W(t) = \prod_{k=1}^{t} \frac{1}{f[A(k)|\underline{A}(k^-), \underline{L}(k^-)]} \qquad (2)$$

where $f[]$ is the conditional probability mass function. In words, $W(t)$ is the inverse of the probability of receiving the observed treatment, conditional on past treatment and covariate history.

Due to the more intuitive nature of MSMs with IPTW over g-computation, and the ability to implement the method using standard statistical routines, there are more examples of this method in the current literature. For example, it has been used to look at questions of concomitant medication in randomised trials [41, 146-150], as well as pharmacoepidemiological studies [42]. It is also possible to implement using standard statistical software packages. As the method of interest in this thesis, further details of IPTW estimation of MSMs will be given in 4.4.1.

4.3.3 G-estimation of structural nested models

Again developed by Robins [39], structural nested models (SNMs) are different from MSMs, in that instead of estimating the effect of treatment on the *marginal* distribution of the potential outcomes, they *condition* on the time-dependent confounder. As outlined by Vandsteelandt and Joffe [151], they deal with the issue of time-dependent confounding by breaking down the effects of treatment into unique incremental effects through time, adjusted for past treatment

and covariate history. There are different kinds of SNMs. For example, Structural Nested Mean Models (SNMMs) [152] are applicable for estimating causal effects on the mean value of an outcome. In simple terms, a SNMM expresses the effect of removing treatment from a given time t onwards on the mean of the subsequent outcome [151], in the subset of patients with the same treatment and covariate history to time $t-1$. For time-to-event outcomes, Structural Nested Distribution Models (SNDMs) such as the Structural Nested Failure Time Model (SNFTM) [153] could be used. These express the effect of removing treatment from a time t onwards, on overall failure time$T > t$.

As a simple example, consider three time points where the continuous outcome Y is measured at time $t = 2$. Exposure A $(0\ or\ 1)$ is assigned at $t = 0$ (defined $A0$) and $t = 1$ (defined $A1$). A covariate L is also measured at these times (with similar notation as for A). Let $Y^{a0\,a1}$ represent the outcome observed if treatment is set to $A0 = a0$, and $A1 = a1$ (note this assumes that the outcome is independent of treatment and covariates at time t=2).

The SNMM would be expressed (as in [151]) as:

$$E[Y^{a0\,a1} - Y^{a0\,0}|A0 = a0, A1 = a1, L0 = l0, L1 = l1] = (\Psi_0^* + \Psi_1^* l_1 + \Psi_2^* a_0)a_1$$

$$E[Y^{a0\,0} - Y^{0\,0}|A0 = a0, L0 = l0] = (\Psi_3^* + \Psi_4^* l_0)a_0 \qquad (3)$$

where the parameters Ψ_0^* … Ψ_4^* are the causal effects to be estimated. E.g. Ψ_0^* is the causal effect of setting $a1$ to 1 vs 0, if $l1$ and $a0 = 0$.

G-estimation is the method by which the set of equations are solved to obtain the estimates for Ψ. Still following the above example, the expressions are transformed to obtain an expression for a function $U^*(\Psi^*)$ of Y such that

$$E[U^*(\Psi^*)|\mathrm{L0}, \mathrm{L1}, \mathrm{A0} = \mathrm{a0}, \mathrm{A1}] = \mathrm{E}[Y^{a0\,0}|\mathrm{L0}, \mathrm{L1}, \mathrm{A0} = \mathrm{a0}, \mathrm{A1}] \qquad (4)$$

Under the assumption of conditional exchangeability (see 4.3.4.3), Ψ^* can be estimated by searching values of Ψ for which the assumption holds (as measured by the "g-test") to obtain the value that is most likely. A confidence interval may be obtained by finding the range of values for which the g-test gives p>0.05 [154], usually done via a grid search [138, 153, 155].

The use of g-estimation of structural nested models in the literature has mostly been in the analysis of survival data. Some recent examples include applications to cancer epidemiology [156, 157], cardiovascular disease [158, 159] and chronic obstructive pulmonary disease (COPD) [160]. The authors of [158] have also written a program for the implementation of g-estimation in Stata [161].

4.3.4 Necessary assumptions

For the methods outline above to provide valid causal estimates, the following assumptions must hold.

4.3.4.1 No interference

A subject's potential outcome under treatment $A = a$ is not affected by the treatment values of other subjects [162]. An example of where this may not hold, is if the exposure were a flu vaccination, and the outcome were contracting the flu virus. In this case, the outcome for an unvaccinated individual may be affected if another subject in the same household were vaccinated.

4.3.4.2 Consistency

Formally, the assumption of consistency implies the following:

$Y_a = Y$ if $A = a$

I.e. The observed outcome under treatment $A = a$ is equal to the potential outcome Y_a. This assumption is usually satisfied by design in a randomised trial, but it may not be automatically true in the observational setting. As explained by Cole [163], informally, the assumption is that, although there are various ways in which treatment a could be assigned, any version of treatment would give the same potential outcome. Careful definition of the exposure can usually ensure this assumption is sufficiently satisfied [35]. An example of where this may not hold, is if exposure was defined as "weight loss". In this situation, weight loss could be achieved by diet, by exercise, or by bariatric surgery. However, it is not clear whether all of these versions of the exposure would result in the same potential outcome.

4.3.4.3 (Conditional) exchangeability

This assumption is also referred to as the assumption of no unmeasured confounders, or ignorability. Mathematically for the example of binary exposure, exchangeability is the assumption that

$$E[Y_a|A=0] \;=\; E\,[Y_a|A=1]$$

In other words, subjects who were not exposed in real life have the same probability of the outcome had they been exposed, as subjects who were exposed in real life.

This assumption can be relaxed to that of conditional exchangeability, which is expressed as follows:

$$E[Y_a|A=0,L] \;=\; E\,[Y_a|A=1,L]$$

I.e. conditional on a set of covariates L (which may be a mixture of fixed baseline and time-dependent confounders), exchangeability holds.

This assumption may be extended to longitudinal data (known as sequential exchangeability or sequential ingorability). In the situation where we measure treatment and covariates at visits $t = 1 \ldots . T$, with outcome Y assessed at time $T + 1$, we assume that:

> "Conditional on treatment history up to visit $t-1$ and the history of all measured covariates up to visit t, the treatment received at visit t is independent of the potential outcomes." ([138] page 1589).

4.3.4.4 Correct model specification

In practice, all methods involve specifying models for treatment assignment, the structural model (model for the outcome) itself, and for g-computation, a model for the covariate history. To obtain valid causal estimates, these models must be correctly specified. For example, modelling age as a linear function when the association is truly quadratic would result in incorrect model specification. For MSM with IPTW, the correct specification of the model for treatment assignment can be relaxed as explained in 4.4.2.3

4.3.4.5 Positivity

In addition to the first four assumptions, the method of IPTW needs the additional assumption of positivity. To define this formally, let $A(t)$ be the treatment at time $t \;=\; 1 \ldots . T$, with $\underline{A}(t^{-})$

and $\underline{L}(t^-)$ as before to be the treatment and covariate history to time $t-1$. Then the positivity assumption can be formally defined as:

$$P(A(t) = a|\underline{A}(t^-),\underline{L}(t^-)) > 0 \ \forall\, a$$

In other words, treatment allocation may not be entirely determined by prior treatment and covariate history. This assumption is necessary for the causal effect to be defined in every subset of the population given the confounders [35].

4.3.5 Method comparison

Daniel et al. usefully provide a clear outline of the strengths and weaknesses of each of the three methods described above [138]. The discussion here will be limited to the advantages and disadvantages most relevant to this thesis. In general, the appropriateness of one method over another would be largely dependent upon the exact research question, and the data available.

In contrast to MSMs with IPTW, neither g-computation nor g-estimation require the positivity assumption to hold. This is advantageous in observational settings where a) strict guidelines or clear contraindications to treatment may cause violations of this assumption, or, b) in smaller sample sizes where the assumption may be violated by chance. In practice, complete violation of this assumption if using IPTW would result in the inability to adjust for the confounding factor that causes the violation. Near violation of the assumption (i.e. a probability of treatment close to 0) results in extremely large weights which can cause bias and instability when applied to the MSM. Simple methods such as weight truncation [164] can reduce the impact of such violations, though may result in increased risk of residual confounding (see 4.4.2.3).

The associations between diabetes treatments and risk of cancer or cardiovascular outcomes are likely to be complex, involving a large number of covariates that are themselves associated with each other. Compared to IPTW of MSMs and g-estimation, g-computation requires the specification of more complex joint distributions, meaning that the chance of model misspecification may be higher [138].

A limitation of g-computation and g-estimation in the context of this thesis is their potential computational cost. The need for Monte Carlo simulation methods and bootstrapping in g-computation, can result in lengthy estimation time. In g-estimation, the use of a searching algorithm could also be computationally intensive. Longitudinal data from the CPRD (see 5.1)

would likely have long follow up, which may need to be split into short time intervals, and coupled with a large sample size (50,000-100,000 for a cohort of incident type 2 diabetes patients), the implementation of the g-methods may be infeasible.

The intuitive nature and ease of implementation (i.e. no need for specialist routines) of MSMs with IPTW, coupled with the likely issue of computational complexity of the g-methods, are the key reasons that MSMs with IPTW have been chosen as the method to be used in this thesis. The following sections will provide a more detailed explanation of the methods, including details of implementation in the statistical package Stata [165].

4.4 MARGINAL STRUCTURAL MODELS

As previously explained in 4.3.2; a marginal strutural model is a model for a counterfatual outcome. Specifically for this thesis, two forms of marginal structural model will be implemented. Firstly, for time to event outcomes, marginal structural Cox proportional hazard (PH) models will be used [144].

As outlined in [145], if we define the survival time T, under treatment history $\underline{a}$ to be the potential outcome $T_{\underline{a}}$,then the potential hazard at time $t = 1 \ldots . T_{\underline{a}}$ is

$$\lambda_{T_{\underline{a}}}(t|V) = \lambda_0(t) \exp\{\beta \underline{a}(t^-) + \boldsymbol{\gamma V}\} \quad [145] \text{ (pg 408)}$$

where V is a vector of baseline covariates, $\underline{a}(t^-)$ is defined as before, and $\lambda_0(t)$ is the baseline hazard.

In words, the marginal structural Cox PH model models the hazard at time t as a function of a baseline hazard λ_0, treatment history to time $t-1$ and baseline covariates only. The baseline hazard therefore represents the hazard for patients who are untreated at all time points, and β is the causal effect of treatment history $\underline{a}$. In this thesis we are interested in the specific case where $\underline{a}$ represents having been treated at all time points.

Secondly, for repeated measures outcomes, a marginal structural generalised estimating equation (GEE) will be used, which is defined as in [166] pg. 1693 (but with slightly different notation) as

$$E\left[Y_{\underline{a}}(t)\middle|, V\right] = \alpha + \beta \underline{a}(t^-) + \gamma t + \boldsymbol{\zeta V}$$

where $Y_{\underline{a}}(t)$is the counterfactual outcome at time t, for a person with treatment history $\underline{a}$.

The marginal structual GEE must be fitted with an independent working correlation matrix to avoid bias, since an exchangable matrix takes information from future time points to inform correlations at previous times [167].

In both of the above models, the effect of treatment is represented by the parameter β. Estimation of β is done via use of IPTW. Interactions between the components of these models can also be included but are not shown here for simplicity.

4.4.1 Inverse probability of treatment weighting of MSMs

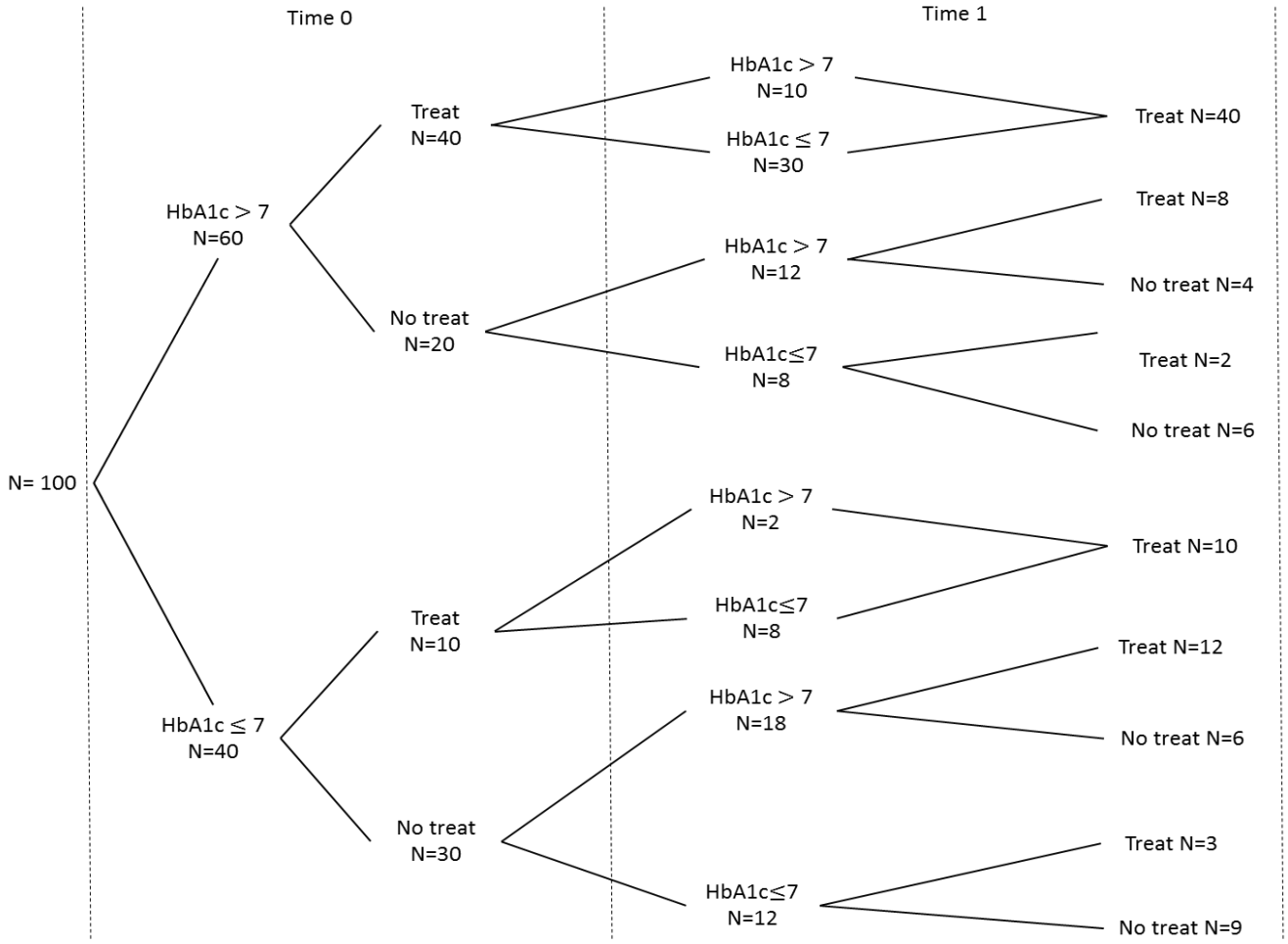
4.4.1.1 A simple example

The following example serves to demonstrate how IPTW removes the association between a time-dependent confounder L and treatment initiation. We consider a simple situation with a single time-dependent confounder and two time points. This example will use the simple unstabilised weight as defined by equation (2) in section 4.3.2.

Consider a population of 100 patients with newly diagnosed type 2 diabetes. At the time of diagnosis (t=0) they all have their HbA1c measured. At this point, the decision to treat with metformin (treated) or recommend a change in diet and lifestyle only (untreated) is made by a clinician, depending on their measured HbA1c. Six months later (t=1), all HbA1cs are measured again, and depending on the value, some patients initially untreated will now commence treatment. All those previously treated remain so. This situation (with numbers treated by strata of HbA1c) is depicted by the tree diagram in Figure 4.2.

.

Figure 4.2 Tree diagram depicting HbA1c and treatment pathways of 100 patients with newly diagnosed t2DM over two time intervals



Although not shown, imagine interest is in the effect of treatment with metformin on an outcome Y (e.g. cardiovascular risk). Assume HbA1c is known to affect cardiovascular risk. It is also predictive of, and affected by treatment. HbA1c therefore satisfies the definition of a time-dependent confounder affected by prior treatment. In order to remove the association between HbA1c and treatment allocation through time, such that risk of the outcome attributable to differences in HbA1c prior to treatment initiation becomes balanced between treatment groups, we calculate IPTW as follows:

Let $A_t = 0 \; or \; 1$ (i.e. untreated or treated at time t), and let L_t denote HbA1c at time t. At time 0, we calculate the probability of observed treatment conditional on HbA1c:

$$P(A_0 = 1 \,|L_0 > 7)) = \frac{40}{60} = \frac{2}{3} \;\; , \quad P(A_0 = 0 \,|L_0 > 7) = \frac{20}{60} = \frac{1}{3}$$

$$P(A_0 = 1 \,|L_0 \leq 7) = \frac{10}{40} = \frac{1}{4} \;\; , \quad P(A_0 = 0 \,|L_0 \leq 7) = \frac{30}{40} = \frac{3}{4}$$

It is clear from this result that treatment allocation is dependent upon HbA1c.

Note that in this example, L_t affects A_t and is assumed to be measured strictly before treatment at time t is determined. Therefore in formula (2), we redefine $\underline{L}(t^-)$ to be the history to time t. Using this we calculate the corresponding weights for subjects within each stratum (by inverting the above probabilities).

$$W_0 = \frac{3}{2} \; if \; A_0 = 1 \; and \; L_0 > 7 \;\; , \quad W_0 = \frac{3}{1} = 3 \; if \; A_0 = 0 \; and \; L_0 > 7$$

$$W_0 = \frac{4}{1} = 4 \; if \; A_0 = 1 \; and \; L_0 \leq 7 \, , \quad W_0 = \frac{4}{3} \; if \; A_0 = 0 \; and \; L_0 \leq 7$$

Moving to the next time point, we calculate the probability of observed treatment conditional on observed treatment history and observed history of HbA1c:

Since if treated at time 0, the patient remains treated, we have that:

$$P(A_1 = 1 \,|A_0 = 1, L_0 > 7, L_1 > 7) = P(A_1 = 1 \,|A_0 = 1, L_0 > 7, L_1 \leq 7 \; = 1$$

$$P(A_1 = 1 \,|A_0 = 1, L_0 \leq 7, L_1 \leq 7) = \; P(A_1 = 1 \,|A_0 = 1, L_0 \leq 7, L_1 > 7) = 1$$

In those untreated at time 0:

$$P(A_1 = 1 \mid A_0 = 0, L_0 > 7, L_1 > 7) = \frac{8}{12} = \frac{2}{3}, \;\; P(A_1 = 1 \mid A_0 = 0, L_0 > 7, L_1 \leq 7) = \frac{2}{8} = \frac{1}{4}$$

$$P(A_1 = 0 \mid A_0 = 0, L_0 > 7, L_1 > 7) = \frac{4}{12} = \frac{1}{3}, \;\; P(A_1 = 0 \mid A_0 = 0, L_0 > 7, L_1 \leq 7) = \frac{6}{8} = \frac{3}{4}$$

The probability of treatment at time 1 is not affected by HbA1c other than the HbA1c at time 1; hence:

$$P(A_1 \mid A_0 = 0, L_0, L_1) = \; P(A_1 \mid A_0 = 0, L_1)$$

It follows that the probabilities for observed treatment at time 1 conditional on Hba1c at time 1 (L_1) for those with $L_0 \leq 7$ are the same as those already calculated for $L_0 > 7$.

The weights for time 1 would be calculated as follows:

$$W_1 = \frac{3}{2} \times 1 \;\; if \; A_0 = 1 \; and \; L_0 > 7 \; , W_1 = 4 \times 1 \;\; if \; A_0 = 1 \; and \; L_0 \leq 7$$

$$W_1 = 3 \times \frac{3}{2} = \frac{9}{2} \;\; if \; A_1 = 1, A_0 = 0 \; , L_0 > 7 \; , \; L_1 > 7$$

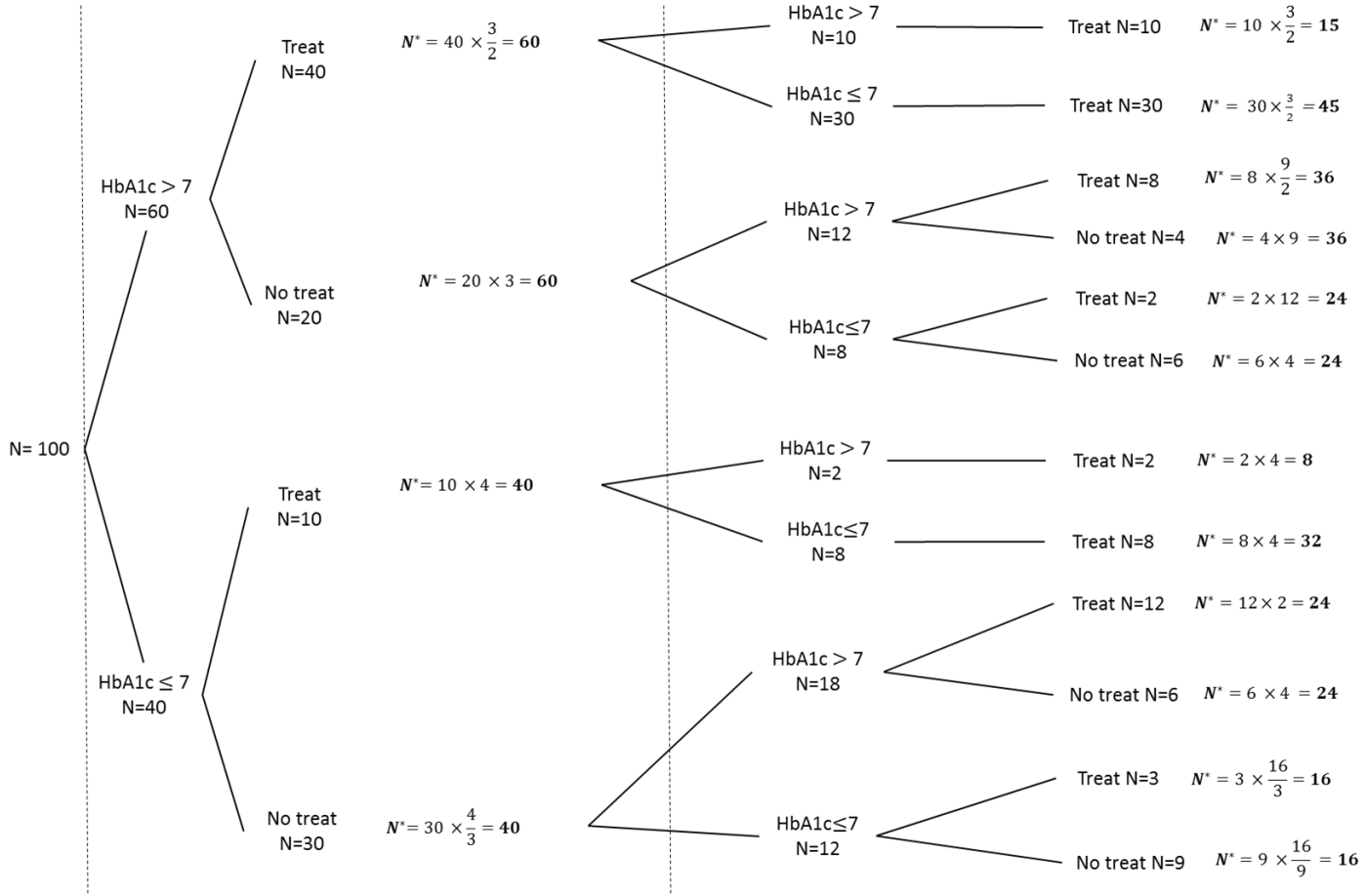
$$W_1 = 3 \times 4 = 12 \;\; if \; A_1 = 1, A_0 = 0 \; , L_0 > 7 \; , \; L_1 \leq 7$$

$$W_1 = 3 \times 3 = 9 \;\; if \; A_1 = 0 \; , A_0 = 0 \; , L_0 > 7 \; , \; L_1 > 7$$

$$etc \; \ldots$$

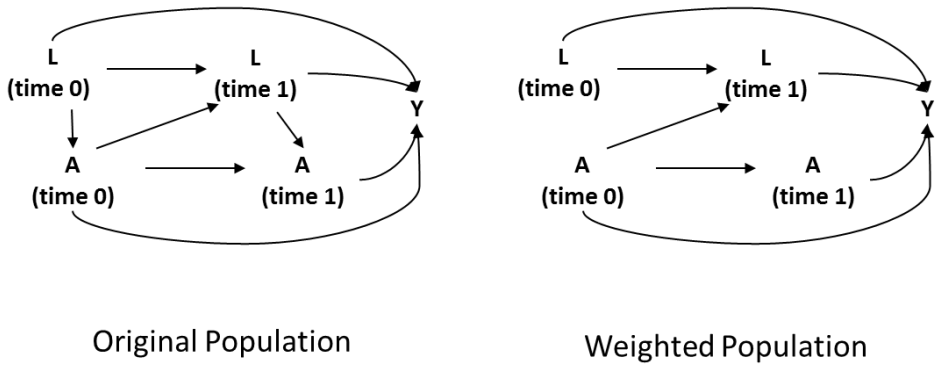
The updated tree diagram in Figure 4.3 shows how the original population N converts to the weighted population N^*. Effectively, we have upweighted the subjects that were under-represented in the original population, i.e. those treated despite a low HbA1c, and those untreated with a high HbA1c. At each time point, in the weighted population, treatment initiation in those untreated at time *t* is no longer dependent upon HbA1c for both $t = 0$ and $t = 1$. However, the distribution of Hba1c at each time point is preserved, as is the effect of treatment at time 0 on HbA1c at time 1. Put simply, the weighting has removed the causal pathway from confounder to treatment without affecting any other associations, as shown in Figure 4.4. The final step would then be to estimate the effect of treatment on outcome in this weighted population without adjustment for HbA1c.

Figure 4.3 Tree diagram depicting HbA1c and treatment pathways of original (N) and inverse probability of treatment weighted population (N)*



p(treat) and p(no treat) defined to be probability of treatment or no treatment respectively, conditional on observed treatment and covariate history.

Figure 4.4 Assumed causal pathways between treatment A, time-dependent confounder L and outcome Y before (left) and after (right) inverse probability of treatment weighting.



4.4.1.2 Weight stabilisation

In practice, using unstabilised weights can result in very large weights. This usually occurs when there are many covariates in the model, and the probability of treatment is very small for some combinations of these covariates. To reduce the impact of this, a stabilised weight is used [37], which has the following formulation (as stated by Fewell [145]):

$$SW(t) = \prod_{k=1}^{t} \frac{f[\,A(k)|\underline{A}(k^-), V]}{f[A(k)|\underline{A}(k^-)\,, \underline{L}(k^-)]}$$

(as per section 4.3.2, $f[]$ is the conditional probability mass function, and here $\underline{L}(k^-)$ is defined as it was originally, as the covariate history to time k-1.)

Informally, the numerator of a stabilised weight is an individual's probability of receiving their treatment at time t, conditional on past treatment history $\underline{A}(k^-)$ and baseline covariates V only (but not conditional on time-dependent confounders).The denominator is the same as for the unstabilised weight (V will be a subset of L). In the simple case, this translates to emulating a population where sequential treatment in patients still off treatment is random conditional on baseline confounders and treatment history.

4.4.1.3 Censoring

In most situations, loss to follow up will occur through time, for reasons such as death , study dropout or administerative censoring in databases. If a subject is censored before the end of follow up, then this can be accounted for via the use of inverse probability of censoring weights (IPCW)[168].

Let $C(k)$ be an indicator for whether a subject is uncensored up to time k, $k = 1...t$, with 0 indicating uncensored, and 1 indicating censored. Then the IPCW at time t, denoted $CW(t)$, is defined as (as in [145])

$$CW(t) = \prod_{k=1}^{t} \frac{f[C(k) \mid \underline{A}(k^{-}), C(k-1)=0, V]}{f[C(k) \mid \underline{A}(k^{-}), C(k-1)=0, \underline{L}(k^{-}), V]}$$

In the presence of censoring, $SW(t)$ must be re-defined so that we estimate the probability of treament conditional on remaining uncensored:

$$SW(t) = \prod_{k=1}^{t} \frac{f[A(k) \mid C(k)=0, \underline{A}(k^{-}), V]}{f[A(k) \mid C(k)=0, \underline{A}(k^{-}), \underline{L}(k^{-}), V]} \quad [145]$$

(where again, $f[]$ is the conditional probability mass function.)

$SW(t)$ and $CW(t)$ can then be multiplied together, in order to obtain a joint inverse probabiltiy weight. This can be thought of as the inverse of the joint probability of observed treatment and remaining uncensored [37]. The resulting weighted population would be interpreted as a population in which there is random treatment allocation with respect to risk of outcome, and no loss to follow up, conditional on baseline covariates and treatment history. The assumption of no unmeasured confounding must extend to the censoring to have this interpretation, meaning it is also necessary to assume that there are no unmeasured common causes of both censoring and outcome.

With the addition of censoring, we now specify that we estimate $Y(t)$ to also be conditional on $C(t) = 0$. That is, we only evaluate the outcome in those remaining uncensored to time t. As specified at the beginning of section 4.4, the risk of the outcome is evaluated with respect to

previous treatment history, i.e. $Y(t)$ depends on $\underline{A}(t^-)$. This has implications in terms of ensuring correct timing when applying the joint weight, since the weight for time t in the model for the MSM must then relate to the probability of $A(t-1)$ and $C(t)$. More formally, by the above notation, the combined weight at time t should be

$$CW(t)\ x\ SW(t-1)$$

where $SW(-1)$ is set to 1.

4.4.1.4 fitting the MSM

Under the assumptions outlined in 4.3.4, the parameters of the MSMs are equivalent to the estimates of the weighted model fitted in the observed data [169], which will be referred to as the "outcome model". For both types of MSM, the practical process of fitting the MSM with IPTW (or joint IPTW/IPCW) is the same apart from the kind of outcome model fitted.

Stata is limited in that time updated weights cannot be applied to a time-varying Cox model. Therefore, the MSM specified at the beginning of 4.4 is approximated by a pooled logistic regression, which is a good approximation providing the probability of the outcome in each separate time interval is small [145, 170].

Formally, for an event Y, define the outcome at each time point $t = 1 \ldots T$, to be $Y(t)$. $Y(t) = 0$ for all intervals until the interval in which the event occurs, at which point $Y(t) = 1$. Then the pooled logistic regression model is defined as:

$$logit\big[P(Y(t)=1|Y(t-1)=0, C(t)=0, \underline{A}(t^-), V)\big] = \alpha(t) + \beta\underline{A}(t^-) + \boldsymbol{\gamma}V \qquad [145]$$

As indicated in the formula, to correctly approximate the hazard ratio that would be obtained from a Cox model, time must be included as a covariate in the logistic regression. How treatment history to time t-1 (denoted $\underline{A}(t^-)$) is modelled may vary. For example, it is often assumed that the effect of treatment history is sufficiently represented by current treatment ($A(t-1)$) only. This approach will estimate an effect of treatment (β) that is constant through time, so has a similar interpreation as a hazard ratio. Specifically, under the assumption that once treated, a subject remains treated, we obtain an effect estimate interpretable as the effect of continuous treatment on risk of outcome, analogous to the estimate obtained from an intention to treat analysis in a clinical trial [171]. Alternatively, if the effect of treatment is cumulative, entering treatment in terms of time since start of treatment may be more appropriate.

Because the stabilisation of the weight means that treatment initiation is balanced conditional on baseline covariates, all variables V included in the numerator of the stabilised weight should be included in the outcome model [143]. This model must also account for the dependence between observations from the same subject which are introduced by the weighting process, and therefore the variance may be estimated by use of a robust variance estimator as described by Hernan, Brumback and Robins [143].

For the MSM for repeated measures, a weighted GEE is used [166]. Time updated weighting is not possible with the Stata command `xtgee` [172], but since an independent working correlation matrx is to be assumed, it can be done using the `glm` command [172].

When using IPTW to estimate an MSM for repeated measures data, historical values of the outcome Y may be included as a time-dependent confounder in the treatment model (i.e. historical values of Y) and the imputation method of last one carried forward (LOCF) can be used where the outcome was not observed for a given interval. In the outcome model, Y is only modelled if observed , meaning the data may be unbalanced.

4.4.2 Practical implementation

Previous literature details the practical step by step process for fitting marginal structural cox models in Stata [145, 170]. In the following sections, the general method as applied in this thesis will be summarised , including the processes implemented to check for and deal with possible violations of assumptions. Further aspects of implementation specific to the analysis in question will be described in the methods of subsequent chapters.

4.4.2.1 Data set up

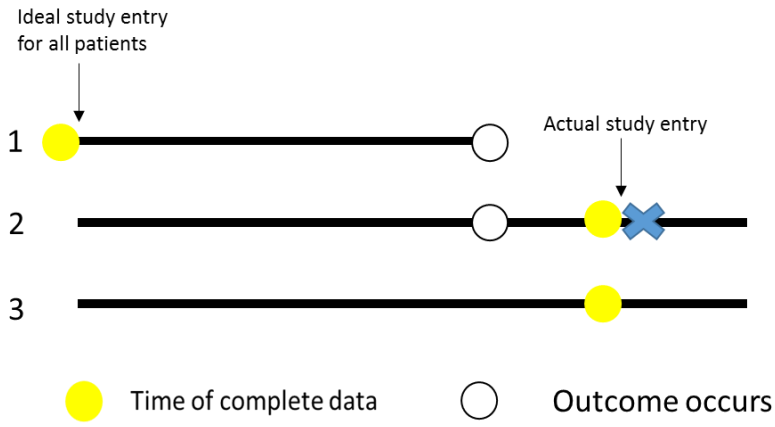
The first step is to split follow up for each subject into discrete time intervals $t = 1 \dots T$. The size of these intervals will be dependent on the question of interest. In each interval, information is required on treatment, covariates, outcome and a censoring indicator $C(t)$.

A weight must be calculated for each interval of follow up, to correctly balance treatment reciept with respect to risk of future outcome. Consequently, follow up can only start once a patient has complete data. “Baseline” will therefore be defined as the **end** of the interval in which complete data occurs (in other words, complete data are obtained in interval t=0). In addition,

patients must still be untreated at the beginning of interval 1, otherwise it is possible that that covariates and outcome have already been influenced by prior treatment.

Looking specifically at the diabetes context and first line treatments, in the ideal situation, all patients would be followed up from time of diabetes diagnosis. For patients for whom study entry is delayed because of incomplete data, it was decided that they would only enter the study once they had complete data, if they would have still been under follow up had they entered the study at the time of diabetes diagnosis. This is demonstrated with an example in Figure 4.5. Here, patient one has complete data from the ideal study entry, i.e. time of diagnosis, and is followed up until their (non fatal) outcome. Patient two has the outcome at the same time as patient one, but does not get complete data until after this time. If they had entered the study at the ideal time they would not still be in follow up by this point, so they do not enter the study. Patient three however, has not had the event prior to their complete data, and therefore enters the study at the time of complete data.

Figure 4.5 example patients to demonstrate rules for study entry if delayed due to incomplete data



For outcomes where patients must have no history of the outcome when they enter the study, this follows automatically. However, it becomes a relevant decision where history of a non fatal outcome at baseline is not an exclusion criteria. The specific situations where this is relevant will be highlighted in later chapters.

Going forwards from study entry, if covariate data are missing for a particular interval, the process of LOCF can be used, which assumes the values remains constant at its last observed value until it is measured again in the data.

If $C(t) = 1$, then the subject is censored from the beginning of this interval. This happens regardless of what else may occur in that interval (e.g. treatment change, change in covariate value, occurrence of outcome).

4.4.2.2 Calcuating IPTW and IPCW

In the example in 4.4.1.1, the probability of treatment at each time point conditional on the history of HbA1c and previous treatment was calculated by hand. In most situations, the number of covariates and number of time intervals means this must be estimated using a parametric model. With the data set up as already described, we can use regression models to estimate these probabilities. For binary treatment, this is a logistic regression model. If there are more than two levels of treatment, multinomial logistic regression can be used [173].

As in the example (see section 4.4.1.1), througout the thesis we make the assumption that once the treatment of interest is initiated, it is continued until end of follow up or a censoring event (which may be a switch to a different treatment). Therefore the probability of treatment in interval t is estimated conditional on being untreated up to the end of interval $t-1$ and uncensored up to the end of interval t. All probabilties of treatment after treatment initiation are equal to one. For survival data, the probability of treatment initiation must also be conditional on the outcome not occuring in interval t. In practical terms, this means fitting the weighting model in those who have not had the event up to the end of interval t.

For the unstabilised weight (or for the denominator of the stabilised weight), the model includes time since baseline, baseline variables (V), and time-dependent variables (L) representing the value of the covariate in the previous interval. This assumes that the baseline and most recent values for L capture full covariate history. However, if this is not the case, then additional variables may be entered to better model covariate history. For example, variables to represent values in further previous intervals t-2, t-3, t-4…. could also be included. For the numerator of the stabilised weight, we again use a logistic/multinomial regression. This time using time since baseline and baseline values of covariates only to predict treatment initiation. This is the "numerator model". In theory, all, or just a subset of basline variables may be used in the stabilisation, but in this thesis all baseline variables will be used.

To estimate the IPCW, we fit separate denominator and numerator models to estimate the probability of being censored in interval t, conditional on remaining uncensored to the end of interval t-1. Since the aim of the IPCW is to balance the censoring process with respect to risk of the outcome, the covariates included in these models should be reflective of those that are predictive of both censoring and outcome. As with the IPTW, time since baseline should be included in both numerator and denominator models. In addition, treatment history should be a covariate in both numerator and denominator because, as explained in 4.4.1.3 we aim to model the joint probability of treament and remaining uncensored. It is likely that there will be multiple censoring mechanisms, e.g. death and study withdrawal. This can be handled by modelling censoring as a categorical variable in a multinomial logistic regression.

The joint weight corresponding to the interval t in the outcome model is then obtained by multiplying by multiplying the IPTW in interval $t-1$ by the IPCW for interval t (where the IPTW for $t=0$ is 1)

4.4.2.3 The bias-variance trade off

The stabilised weights at each time t should be approximately normally distributed with an expected value of one [174]. Larger means can be an indication of violations of the positivity assumption, since they may indicate that some individuals have a covariate history that strongly predicts their treatment history. Large, or "extreme" weights may also be an indication of poorly specified treatment models [169]. Using extreme weights in the outcome model will reduce the precision of the treatment effect estimate, and may cause bias. On the other hand, methods to reduce extreme weights can re-introduce bias from confounding. This is known as the bias-variance trade off [169]. As such, fitting the treatment and censoring models must be done with the aim of finding the best model to balance residual confounding with reduced variance or bias from positivity violations.

As suggested by Cole and Hernan [169] examination of the weight distribution is a good way to identify possible violations of assumptions during the model fitting process. As they explain, although the distribution should be ideally examined at each time point, with long follow up and many time intervals, this may not be practical, and so the distribution of all the weights can be examined instead.

A common approach to deal with near positivity violations is weight truncation. Simply, extreme values are truncated in order to reduce their impact on the overall distribution of the weights.

This approach was advocated by Cole and Hernan [169] as a good approach to balance the bias-variance trade off, since gradual truncation can provide information on the extent to which variance improves, but potential bias may be introduced (as indicated by changes in effect estimate).

In order to calculate weights that correctly balance treatment with respect to risk of outcome, it is important that the treatment model is correctly specified [175] , most importantly for variables that are predictive of outcome. This necessitates having the correct variables in the model, and that the functional forms of the associations between covariates and log odds of treatment/censoring are correctly specified. The issue of modelling the correct functional form is also relevant to the bias-variance trade off. Also demonstrated by Cole and Hernan [169], fine modelling of covariates (in their example, using a large number of categorisations) will increase control of confounding but may increase variance due to positivity violations. They recommended that multiple forms of covariates are tested to compare how variance and possible bias of the estimate of interest are impacted by fine vs coarse modelling.

One of the key things to consider when deciding how finely to model the association between the variable and treatment is how much residual confounding might be expected if the functional form for a particular variable is not precisely specified in the treatment model. It is most important to have the correct form over intervals/values for which the risk of outcome changes. If the risk of outcome is roughly the same for all values of the covariate across a particular range, then a misspecification of the association between treatment and the covariate over this range is unlikely to result in serious residual confounding.

In terms of variable selection, Lefebvre et al recommend that variables should be selected based primarily on their association with the outcome. This is because including variables that are strongly associated with treatment and not outcome will increase variance and cause bias in the causal effect estimate [175]. Variable selection may be based on a priori clinical knowledge, or by looking at the observed associations between potential confounders and outcome within the data.

4.4.2.4 Planned approach to obtaining treatment and censoring models

For this thesis, the approach to obtaining treatment and censoring models that adequately balance the bias-variance trade was a relatively subjective process guided by the recommendations described above. A priori discussions with clinicians established a set of

confounders and risk factors for the outcome in question to be included in the models, though the exact form of these variables was not pre-specified. Additional covariates were also considered for inclusion if initial analysis of association with outcome suggested they might be important.

Two model specifications were used as a starting point to look at the impact of potential model misspecification:

1. continuous variables were fitted using natural cubic splines (spline form)
2. all covariates fitted as categorical variables (categorical form)

Time was always entered into the model as a continuous covariate, represented by a natural cubic spline, though the exact form of the spline was allowed to vary between specifications 1 and 2. Further details of how these models were obtained and developed for the specific analysis are described in 7.2.7.2 and 8.2.5.2.1.

The distribution of the weights for the different model specifications was then examined to get an indication of whether different specifications had a large impact on the size of the weights. For each model specification, weights were truncated to see how this affected the overall mean of the weights. Truncation at the 1st and 99th percentile was considered first, though a range of more severe or more lenient truncations were also considered.

4.4.2.5 Correct specification of the MSM

As with the treatment and censoring models, the outcome model must be correctly specified. The correct functional forms of time and baseline covariates may differ between the treatment, censoring and outcome models. For example, time may have a linear association with probability of treatment, but a non-linear association with risk of outcome.

When fitting the outcome model, multiple forms for covariates were also investigated. As with treatment and censoring models, the two initial specifications were based on:

1. Natural cubic splines for continuous covariates
2. Categorisations for all covariates.

For practicality, the same spline parameterisations were used in all models if they appeared to model all associations with sufficient accuracy. Further details as to how these models were

established are given in later chapters. Different forms of treatment were also investigated to ensure treatment history was sufficiently captured.

4.5 DYNAMIC MARGINAL STRUCTURAL MODELS

4.5.1 Introduction

A static treatment strategy (or regime) is one that is decided from time 0 and is fixed. Examples of static regimes could be: always treat, never treat, or treat every other month. To this point, the focus has been on using inverse probability of treatment weighting to remove the dependency of treatment on time-dependent confounders to allow the estimation of the causal effect of static treatment strategies.

However, treatment strategies that include dependency on time-dependent covariates may also be of interest and are frequently used in clinical settings. Type 2 diabetes medications are initiated and intensified in response to blood glucose measures, for example, the current UK clinical guidelines advocate the treatment strategy "initiate medication if HbA1c raises above the threshold of 6.5%" [13]. Such a treatment strategy is known as a dynamic treatment strategy (or regime), and comparison of such strategies may enable the identification of the optimal strategy to improve long term outcomes for patients.

Dynamic regimes are rarely compared in clinical trial settings, most likely because if the interest were in comparing many different thresholds, a very large number of participants would be required, and the trial would be extremely expensive. In the observational setting, variation in and compliance with strategies will likely be dependent on factors that affect risk of outcome, and so time-dependent confounding will be an issue.

The three methods described in 4.3 have all been extended to allow estimation of the causal effects of dynamic regimes [176-180]. Since MSMs are the focus of this thesis, the description of methods here will be limited to a MSM approach. Two extensions of MSMs have been suggested to compare dynamic treatment regimes: History adjusted marginal structural models (HAMSMs)[178], and dynamic marginal structural models (dMSMs) [179, 181], the latter being the approach of interest going forward. Some examples of the use of dynamic marginal structural models in existing literature include comparing dynamic treatment strategies for HIV [182], schizophrenia [183] and cancer[184].

4.5.2 The basic idea

Dynamic marginal structural models aim to emulate the randomised trial where patients are randomised to m interventions, which in some way depend upon the value x of a time-dependent variable L. For the purposes of the thesis, we assume an intervention of the form "treat when $\boldsymbol{l}$ **exceeds** $\boldsymbol{x}$", for m different values of x. x may also be referred to as the "treatment threshold". Specifically, we take L to be a patient's HbA1c.

Different patients in observational data will be compliant with different regimes. At the start of follow up, all patients are compliant with all possible regimes (or we restrict the population so that this is the case). With time, patients will be compliant with regime x until they are either treated with $l < x$, or remain untreated with $l > x$, at which point they become noncompliant.

The basic intuition is to expand the data such that each patient's follow up is duplicated for each regime. For each regime, a patient is followed up until they become noncompliant with the regime, at which point they are censored. We estimate inverse probability of censoring weights to produce a weighted population in which noncompliance with a particular regime is not dependent upon subsequent risk of outcome, conditional on covariate history. Modelling the association between treatment regime and outcome in this weighted population will allow, under the same assumptions as for static regimes, the causal effect of the regime on risk of outcome to be estimated.

4.5.3 Formal notation of method

The notation and formulation of the models given here broadly follows that used by Ewings et al [182].

4.5.3.1 The dynamic marginal structural Cox model

As in section 4.4, consider a time-to-event outcome Y with survival time T, with $Y(t)$ (=0 or 1 to mean no or yes respectively) representing whether the outcome has occurred by time t. Here, times are discrete time intervals, such that t is meant to represent the interval $(t-1, t]$.

Define the different possible treatment regimens by the thresholds at which treatment should be initiated, as $x = x_1, x_2 \ldots . x_m$. The counterfactual survival time under regime x is denoted T_x , where T_x will be observed for a subset of x for each subject.

Then, the dynamic marginal structural cox model has the form

$$\lambda_{T_x}(t|\underline{V}) = \lambda_0(t)\exp\{\beta g(x,t) + \boldsymbol{\gamma} V\} \qquad [182]$$

where $\lambda_0(t)$ and $V$are defined as in 4.4, and $g(x,t)$ is some function of the strategies, which usually includes an interaction with time.

4.5.3.2 The censoring process

To indicate whether a subject is still compliant with regime x (and so uncensored) at time t, define $C_x(t) = 0$ to mean they are compliant, and $C_x(t) = 1$ to indicate they are no longer compliant. Further to this, for each regime x , an indicator variable is needed to know whether the observed value HbA1c, has exceeded the threshold x by time t. We denote this indicator $Q_x(t)$ to be equal to 1 if HbA1c exceeds the threshold in interval t, and 0 if HbA1c remains below the threshold up to the end of interval t. $C_x(t)$ is then determined by treatment at time t (denoted $A(t)$ as before), $Q_x(t-1)$, and $Y(t)$.

Specifically

$C_x(t) = 0$ if for all times $k = 1 \ldots t$, when $Q_x(k-1) = 0$, $A(k) = 0$ and $Y(t) = 0$,

or when $Q_x(k-1) = 1, A(k) = 1$ and $Y(t) = 0$ [182]

It will be assumed in all analyses, that once treatment is initiated, it should be continued regardless of further changes in HbA1c. Therefore if treatment is initiated in line with regime x, then the patient is compliant with x for the rest of follow up. It follows therefore, that if $A(t) = 1$ and $C_x(t) = 0$, then $C_x(s) = 0$ for all $s > t$. Additionally, once a patient is censored they remain censored even if they later become compliant with the strategy.

4.5.3.3 Inverse probability weighting

The censoring process formulated above is likely to be dependent upon time-varying and baseline factors that also affect risk of outcome. As before, the vector of baseline covariates is denoted by V and the history of a covariate from time $1 \ldots t-1$ by $\underline{L}(t^-)$. Under the same

assumptions of no unmeasured confounding (in this case for censoring mechanism and outcome), exchangeability, consistency, positivity, and correct model specification; the informative censoring process can be adjusted for by using inverse probability weights, defined similarly to in 4.4.1.3; as:

$$W_x(t) = \prod_{k=1}^{t} \frac{1}{P(C_x(k)=0|C_x(k-1)=0,\underline{L}(k^-),V,Y(k)=0)} \text{ if } C_x(k) = 0 \text{ and 0 otherwise.}$$

where $W_x(t)$ is weight at time t for an individual uncensored to t on regime x.

Note that in contrast to 4.4.1.3, treatment history $\underline{A}(t^-)$ is not included. Additionally, only time-to-event outcomes are considered here, so the condition that the outcome has not occurred up to time is included in the formal definition. In fact, since $C(t)$ is a function of A, Y and Q, where Q is dependent upon L, this weight is equivalent to the inverse of the probability of treatment at time t, conditional on treatment and covariate history:

$$W_x(t) = \prod_{k=1}^{t} \frac{1}{P(A(k)|\underline{A}(k^-),\underline{L}(k^-),\underline{V},Y(k)=0)} \text{ if } C_x(k) = 0 \text{ and 0 otherwise.}$$

To avoid confusion, despite the equivalence to the IPTW, in the dynamic context these will be referred to just as "inverse probability weights". Formulations of stabilised weights have been proposed by Cain et al [181] , however the authors report that the stabilisation cannot guarantee less variable estimates than unstabilised weights. Additionally, stabilisation is not well developed for use with grace periods (see section 4.5.5), and with sensible regime choice, extreme weights are less of an issue. Consequently, for simplicity, only unstabilised weights are presented and implemented in this thesis.

4.5.3.4 Dealing with additional loss to follow up

Other informative censoring (loss to follow up) can be dealt with by separate IPCW. This is done in the same way as described in 4.4.1.3, but for clarity is presented again within the dynamic framework. Define $D(t)$ to be 0 if a patient is not lost to follow up in interval t, and 1 if the patient is lost to follow up in interval t. Then we calculate the IPCW as

$$DW(t) = \prod_{k=1}^{t} \frac{1}{P(D(k) = 0|\underline{A}(k^-), D(k-1) = 0, \underline{L}(k^-), V)}$$

and the formula for $W_x(t)$ becomes

$$W_x(t) = \prod_{k=1}^{t} \frac{1}{P(A(k)|A(k^-),L(k^-),D(k)=0,\underline{V},Y(k)=0)} \text{ if } C_x(k) = 0 \text{ and 0 otherwise.}$$

4.5.4 Practical implementation

4.5.4.1 Defining plausible strategies for comparison

It is important to consider in advance, the range of x for which the data are compatible. Therefore the number of subjects compliant with various strategies should be examined in advance, and strategies with low compliance excluded.

4.5.4.2 Calculating weights

As was the case for comparison of static regimes, data for each subject are again split into discrete time intervals $t = 1 \ldots . T$. Calculation of the weight at time t are independent of the specific strategy x, so unstabilised inverse probability weights can be calculated before any expansion takes place[181, 182]. The process is the same as described in 4.4.2.

4.5.4.3 Data expansion and censoring

For each of m values of x (the treatment initiation thresholds being compared), a subject's data must be replicated. In the expanded data, the indicator for $C_x(t)$ is created so that subjects are censored once they become noncompliant with regime x. Therefore, in contrast to censoring by loss to follow up, the subject is not censored until the **end** of the first interval in which $C_x(t) = 1$. In other words, we evaluate the risk of outcome for x in those compliant with x up to the end of the previous interval. This avoids bias that could be induced by censoring events that occur in the same interval in which a patient becomes noncompliant with a particular regime, that would not have been censored if they had remained compliant in that interval [185]. For example, bias could be induced if occurrence of the outcome affects ability to comply. This assumes, in the same way as for the standard MSM, that the risk of outcome is not affected by treatment in the same interval.

4.5.4.4 Estimation of the dynamic MSM

To estimate the parameters of the dynamic MSM presented in 4.5.3.1; we again approximate a Cox PH model using pooled logistic regression [170]. In contrast to the standard MSM, this model includes an interaction between time and treatment regime. This is important because unless the risk of outcome is the same between the regimes, any difference between regimes cannot be constant through time, since some follow up is compatible with multiple regimes. The general model used is of the form

$$logit\big[P(Y(t)=1|Y(t-1)=0, C_x(t-1)=0, x, \underline{V}\big)\,]$$

$$= \alpha f(t-1) + \beta g(x) + \boldsymbol{\gamma}\underline{V} + \delta g(x) f(t-1)$$

where f and g are functions of time and regime respectively. As with the standard MSM, the weight applied to time t must be that relating to probability of treatment at time $t-1$ [182], and if additional IPCW are used to account for loss to follow up, we additionally condition on remaining uncensored in that interval. More formally, we evaluate

$$P(Y(t)=1|Y(t-1)=0, C_x(t-1)=0, D(t)=0, x, \underline{V})$$

This model is solved in the weighted expanded population, in order to specify a functional form for x [181].Confidence intervals for the effect of regime over time can be estimated using non-parametric bootstrapping [186].

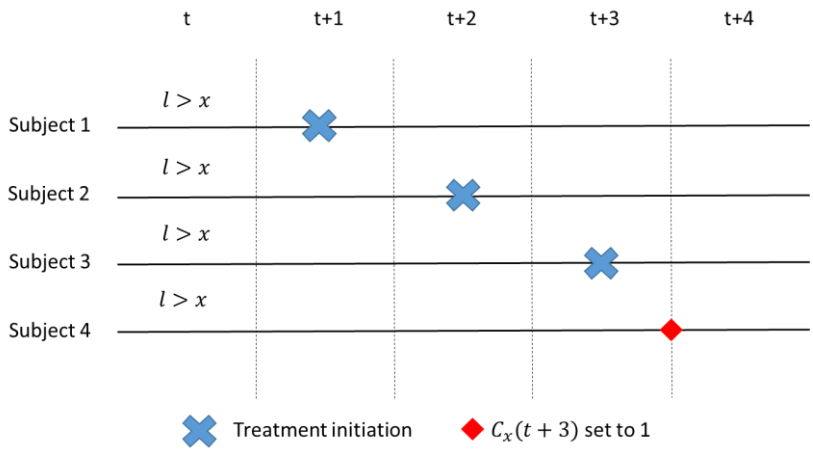
4.5.5 Grace periods

To this point, compliance with a regime when treatment is indicated has been defined by initiation in the interval $t+1$ if $Q_x(t) = 1$. In real life settings such a short period to allow treatment initiation may be impractical. It may be more realistic, and increase numbers complying to given regimes, to allow initiation within a longer period. These extended initiation periods are known as grace periods [181]. Under a grace period, the regimes are no longer fully identified. Applying a grace period p would alter the regime to be of the form: "treat within p intervals of the time point at which l first exceeds x". To demonstrate this, Figure 4.6 shows four subjects observed over five consecutive time intervals, where in the first time interval t, all four subjects have their first measurement of l above the treatment threshold x.

Assume that we have a grace period $p = 3$. Subject 1 initiates treatment in the first interval following a measure of $l > x$, and is therefore compliant with regime x in the same way that

compliance was originally defined (which could be seen as applying a grace period of $p = 1$ [185]). Subjects 2 and 3 also initiate within three intervals, and therefore they are also compliant. Subject 4 however, gets to the end of t+3 and has not yet initiated treatment, therefore $C(t+3)$ is set to 1, and follow up for subject 4 in terms of risk of outcome ends at the end of interval $t+3$.

Figure 4.6 Graphical representation of how compliance with regime $\boldsymbol{x}$ is assessed with a grace period $\boldsymbol{p = 3}$.



The use of grace periods requires modifications to the inverse probability weights, because it is impossible to be noncompliant with the regime until the last interval of the grace period. Cain et al [181] have suggested two possible approaches to how the weighting is adjusted during the grace period. The simplest approach (the one used in this thesis) is if the patient is uncensored at the beginning of the grace period, to set the probability of remaining uncensored ($P(C_x(t) = 0)$) to be one in all intervals within the grace period other than the final interval. This is because once the grace period starts, the subject cannot become noncompliant before the end of the final interval. In practice, this means that we upweight the subjects that initiate in the final interval of the grace period to account for those who are censored at the end of the grace period because they do not initiate treatment. This approach makes the assumption that the probability of initiation will not be uniform across the grace period. More specifically, it re-defines the regime to be: "treat in the pth interval following that when HbA1c first exceeds x% if treatment has not already been initiated within the $p-1$ intervals following that when HbA1c first exceeds x%".

Formally, if we take time q_x to be a time interval in which $Q_x(q_x - 1) = 0$ and $Q_x(q_x) = 1$ then for a grace period of length p:

$$W_x(t) = \prod_{k=1}^{t} \frac{1}{P(C_x(k)|C_x(k-1)=0, L(k^-), V, Y(k)=0)} \quad \text{if } C_x(k) = 0 \text{ and 0 otherwise}$$

$$\text{where } P(C_x(k) = 0 | C_x(k-1) = 0) \equiv 1 \; for \; q_x + 1 \leq k < q_x + p$$

$$\text{and } P(C_x(k) = 0) = (P\big(A(k) = 0 \big| \underline{A}(k^-), \underline{L}(k^-), \underline{V}, Y(k) = 0\big) \text{ for } k \leq q_x$$

$$\text{and } P(C_x(k) = 0) = (P\big(A(k) \big| \underline{A}(k^-), \underline{L}(k^-), \underline{V}, Y(k) = 0\big) \; for \; k = q_x + p$$

The initial weight estimation is the same as described in 4.5.4.2. After data expansion, the weights can be adjusted as explained above. The dynamic MSM is then fitted in the same way as described in section 4.5.4.4.

5 DATA SOURCE AND COHORT IDENTIFICATION

The source population of interest was adult patients with newly diagnosed type 2 diabetes, including those controlling their diabetes with diet and lifestyle measures. This chapter describes the data resource used and the process implemented to identify a study sample to represent this population.

5.1 THE CLINICAL PRACTICE RESEARCH DATALINK

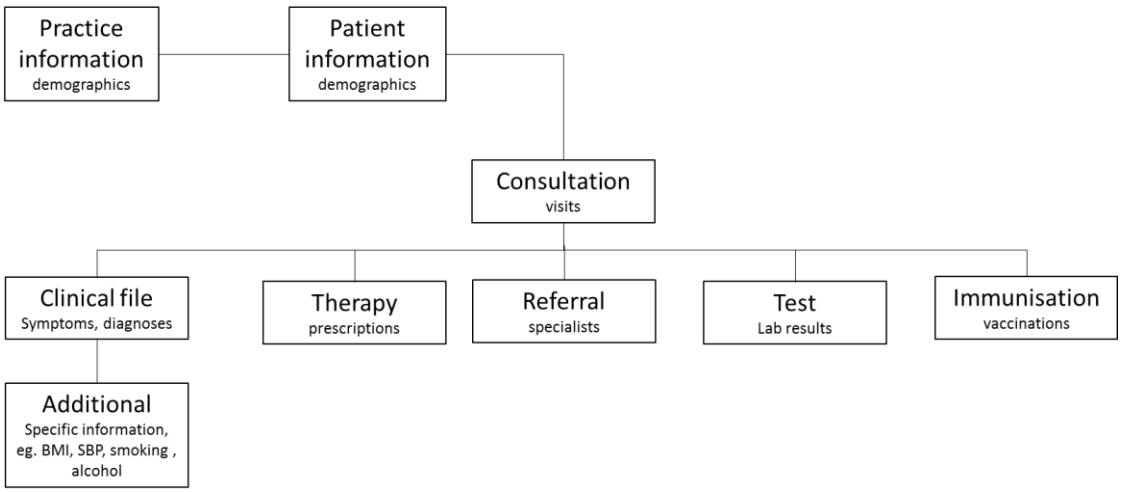
The Clinical Practice Research Datalink (CPRD) (formerly known as the General Practice Research Database (GPRD)) is a database containing general practice (GP) records from 691 UK practices as of the January 2016 extract, with over 14 million patients contributing records [187]. GPs are free at the point of use, and provide the first point of contact for non-urgent healthcare. They provide multiple services, including care for ongoing chronic disease, treating minor illnesses, and providing the necessary referrals to secondary care in hospitals or specialist clinics. The CPRD contains all available data from primary care consultations, including clinical diagnoses, prescription information, test results (including HbA1c), and other clinical information such as BMI, blood pressure, smoking status etc. Information from secondary care, for example, diagnoses of conditions, are also fed back to the GP and should in theory also end up in the record. However, some information may be lost if only typed into free text data fields, as these are not available for research use currently. Many validation studies have shown CPRD primary care data to have good validity for a wide range of diagnoses [188]. Those in the database have been shown to be broadly representative of the UK population in terms of age, gender and ethnicity [189]. In 2005, incentivised recording and monitoring for certain chronic conditions including diabetes was introduced, meaning that the data quality and completeness from this time is much improved [190].

Briefly, the different aspects of CPRD (prescriptions, test results, clinical consultations etc.) are linked by unique patient identification numbers, and each have an event date by which to order events. For individual patients, entry and exit times from the database can be identified by date of registration and date of transfer out or date of death (if entered by the GP). For patients that have neither left the practice nor died, each patient can be linked to specific practice information to detail the date of last data collection for that practice. This allows reasons and timings of

censoring to be identified. The data structure, as previously presented by [189] is outlined in Figure 5.1.

CPRD can be linked to other data sources such as Hospital Episode Statistics, cancer registries and national death records and to practice level measures of deprivation, though the addition of linked data was not within the scope of this thesis.

Figure 5.1 Data structure for CPRD, recreated from Herrett et al. figure 2 [189]



5.2 COHORT IDENTIFICATION

5.2.1 Biobank algorithm overview

Pre-existing code lists for patients with diabetes have been developed for CPRD based on an algorithm originally developed for the UK Biobank data [191]. These lists were developed in order to identify patients with prevalent type I and II diabetes.

Briefly, the UK Biobank algorithm extracts patients based on an **initial diagnosis** code list, which will be referred to as primary diagnosis codes. The algorithm then aims to confirm diagnoses of diabetes by assessing the additional presence of

a) Oral antidiabetic medication

or

b) Codes relating to the care and management of diabetes within a patient's medical history, termed "process of care" codes, such as "diabetic monitoring" and "diabetic annual review".

The code "diabetic on diet only" is classified by BioBank as a process of care code, and not a diagnostic code. For the purposes of this research, "diabetic on diet only" was moved into the primary diagnosis code list, because initial data exploration suggested that this code was often used in CPRD to denote a diagnosis with no further diagnosis code present. Following the Biobank algorithm may have led to the exclusion of a lot of pre-medication follow up. All other process of care codes were used as in the original algorithm. All code lists used are presented in appendix 7.

For some patients, information is also required on ethnicity, BMI, and age to judge whether it is likely that the patient has type 1 or type 2 diabetes. For the application to this thesis, the dependence upon ethnicity was removed from the algorithm, and this part of the algorithm was based on BMI and age only. The reason for this was to improve the sample size by not excluding a large proportion of patients who did not have ethnicity recorded. The implementation of the algorithm with the modifications described above is described in the following section.

5.2.2 Algorithm implementation

The flow chart in Figure 5.2 shows a simplification of the overall algorithm. In an initial step, primary diagnosis codes in a patient's history were sorted to place the patient into their most likely category of the following: Definite Type 1, Definite Type 2, Probable Type 1, Probable Type 2, Possible Type 1, Possible Type 2, Vague codes (referring to codes that are unclear), Genetic diabetes, and Other Types (secondary, gestational, not diabetes, resolved diabetes).

These categories were allocated in a hierarchical way; with definite codes having priority over probable codes within the same type, and type 1 codes taking priority over type 2 codes overall. For example, a definite type 2 code was given priority over a possible type 2 code, but regardless of other codes, if a patient had a code suggesting type 1 diabetes, they were initially classified with that. Where patients had conflicting codes in relation to "other types", rules based on the timing of codes were implemented to pick the most appropriate code to take forward to the next stage. For example, a patient with confirmed gestational diabetes after a previous possible type 2 code was classed as having gestational diabetes. The patients were then given an initial "date of onset" corresponding to the first occurrence of a code from the category they had been assigned. This initial process corresponds to stage 1 of the flow chart in Figure 5.2, where

definite, probable, and possible for each type are combined for simplicity. Codes considered to be “vague” were included with the possible type 2 diabetes. Genetic diabetes was combined with “other types”, and patients assigned any of these codes were excluded. A more detailed flow chart for this initial stage is presented in appendix 8.

The second stages of the algorithm also required information on whether the patient had ever been prescribed insulin, metformin, or other OADs (see appendix 9), and also:

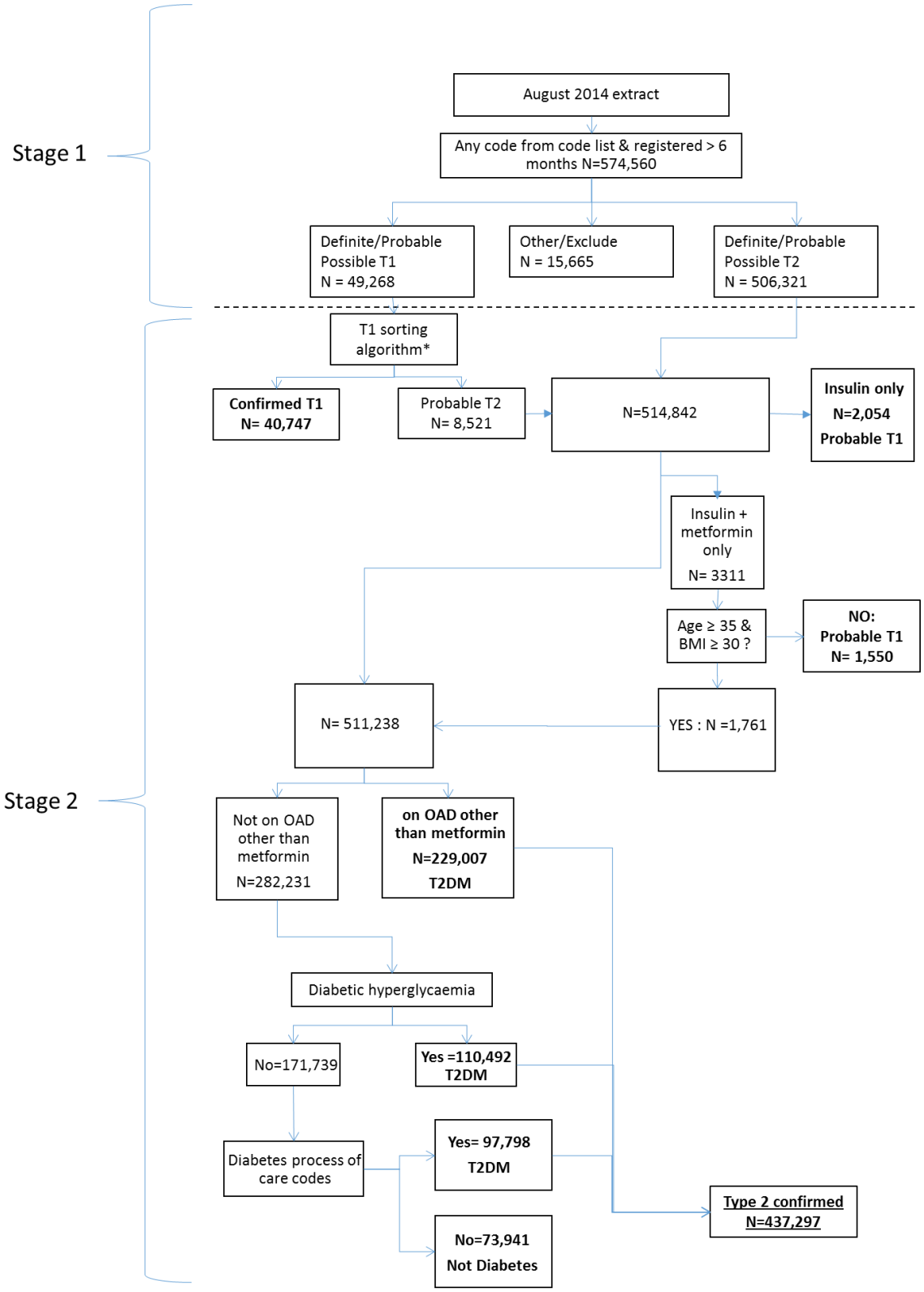
- The age of the patient at their date of onset.
- The BMI of the patients closest to their date of onset (see appendix 9).
- Presence of hyperglycaemia (defined as HbA1c > 6.5% or fasting glucose >7.5 mmol/L) at the closest time to date of onset available (see appendix 9).
- Process of care codes to indicate that the patient was receiving ongoing care for DM.

Two further sorting algorithms were then applied to the data. Firstly, those patients initially identified as definite, probable or possible type 1 were sorted and either kept as a type 1, or re-entered as possible type 2. In brief, patients entering this algorithm were immediately reclassified as possible type 2 if they had never been prescribed insulin; if they had been on an oral anti-diabetes drug other than metformin for more than 6 months; or if they were ever on insulin and metformin, or were overweight and over the age of 35 at the time of the first diagnosis code. All remaining patients were classified as having type 1 diabetes. The re classified patients were put back into the next stage of the algorithm. A flow chart to show how patients were sorted through this algorithm is given in appendix 8.

Secondly those initially sorted as definite, probable, possible type 2, vague, or patients coming back into the pool of possible patients from the type 1 algorithm were sorted by a second algorithm into type 2, probably type 1, and not diabetes. This process corresponds to stage 2 of Figure 5.2, with a more detailed flow chart of this step given in appendix 8.

The 437,297 patients remaining as type 2 were kept as a cohort of patients with both incident and prevalent type 2 diabetes.

Figure 5.2 Flow chart to show numbers identified as T2DM patients from August 2014 CPRD extract



*See appendix 8 for details

5.2.3 Defining onset, start of follow up, and final incident diabetes cohort

Once the initial cohort of 437,297 patients with T2DM was identified, a more specific date of diabetes onset was defined. This process was not part of the original Biobank algorithm, but was implemented to obtain a cohort of patients that could (with greater certainty) be followed up from time of diabetes diagnosis. In consideration of the order in which the algorithm priorities information, but also to maintain simplicity, the patient's date of diabetes onset was taken to be the earliest of:

a) their earliest of a possible/probable/definite T2 diabetes code
b) the earliest process of care (POC) code
c) the earliest medication with either metformin or an OAD.

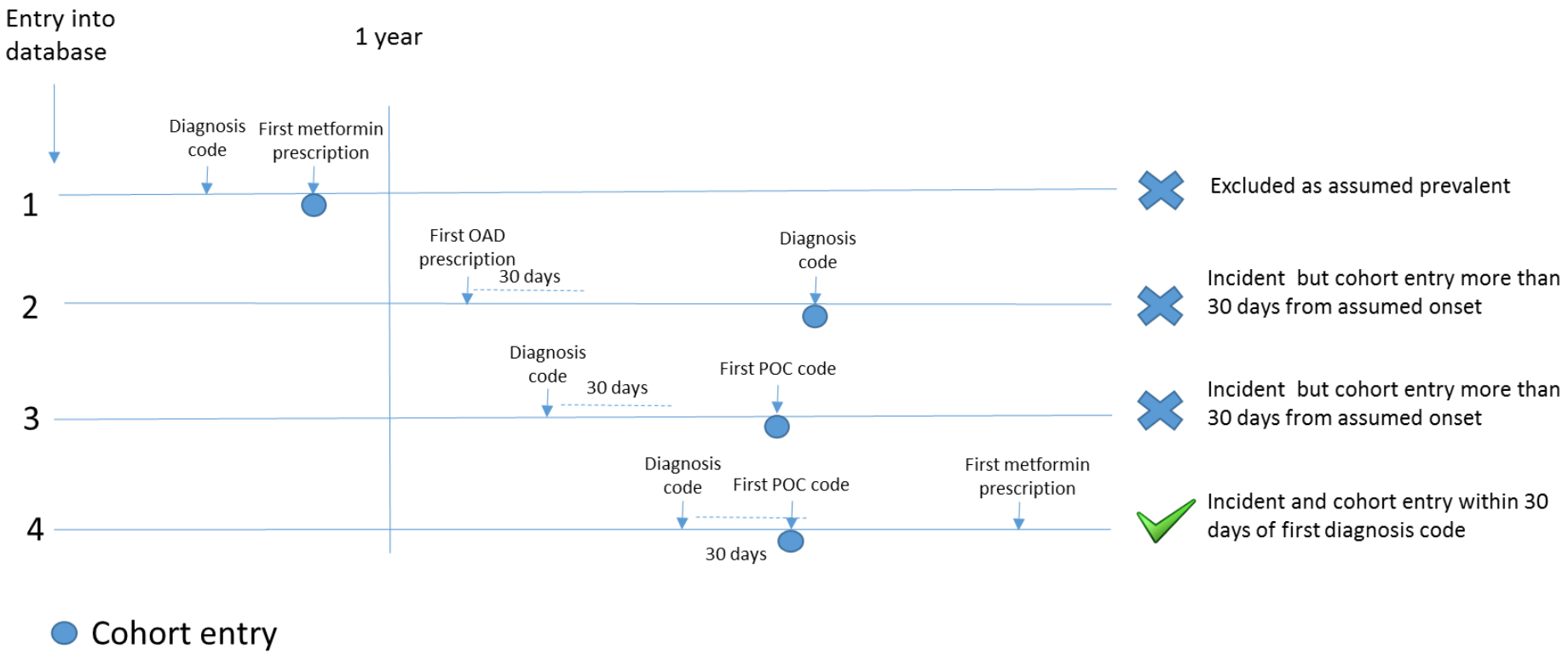
Firstly, the data were further cleaned to ensure all patients within the final cohort had were of a reasonable age at time of diagnosis. It was decided to exclude patients aged < 30 or > 90 years at diabetes diagnosis since the number of incidence cases under 30 was expected to be low [16, 192] and, very different treatment strategies would be expected to apply to these ages, particularly for patients aged > 90 , due to increased frailty. This reduced the available sample size to 418,473.

For a patient to be considered an "incident" rather than a prevalent case, it was required that there were at least 12 months of history prior to their onset date with no diabetes related codes. This was because when patients enter the database (are registered with the GP), previous history of pre-existing conditions may be entered into the database, and so using codes within the first year is likely to pick up prevalent cases of diabetes [193]. As an example, patient 1 in Figure 5.3 would not satisfy this requirement. Since multiple codes define the inclusion of a patient into the cohort, it was important to ensure that a patient did not begin follow up until the time when the minimum number of codes that defined inclusion had occurred. Alongside the diagnosis code, the three points in the Biobank algorithm that defined final inclusion were a prescription for an oral anti-diabetes drug other than metformin, a test indicating hyperglycaemia, or a process of care code. To ensure a more certain diagnosis of diabetes, the decision was made to require a process of care code in addition to the record of hyperglycaemia (resulting in the loss of about 6,000 of the 110,492 identified by the Biobank Algorithm (Figure 5.2). As shown in Figure 5.3 by the circle, the date of cohort entry (follow up start) was defined as the latest of

a) the first T2 diabetes code, and
b) the first process of care code OR first medication with an OAD other than metformin.

To be sure that the cohort was still a cohort of incident cases of T2DM, the follow up start date and the date of onset were further required to be within 30 days of each other. Patients satisfying these criteria were those taken as the final cohort for study. In Figure 5.3 only patient 4 satisfies this final assumption. Finally, all patients were required to have had no history of cancer at the time of diabetes diagnosis. This would be necessary when examining cancer incidence as an outcome, but was applied for all analyses to avoid including patients whose future risk of diabetes, and complications from diabetes, may have been modified by cancer in a way that cannot be clearly controlled for. These additional requirements resulted in a final incident cohort of 98,080 patients.

Figure 5.3 Example patients to demonstrate inclusion and exclusion rules to obtain the final cohort of patients with incident type 2 diabetes.



6 TREATMENT PATTERNS AND FREQUENCY OF COVARIATE MEASUREMENT

6.1 MOTIVATION AND AIMS

It is important in advance to have a well-defined causal question and exposure[194], and more broadly to know it would be feasible to answer this question with the available data. As outlined in 1.3, there are a limited number of existing applications of IPTW of MSMs in both the diabetes context and in routinely collected primary care data and particular aspects of this setting may present issues in terms of the feasibility of the method. Two key areas for concern are a) possible violations of the positivity assumption (4.3.4.5) due to strict guidelines for diabetes management, and b) visit frequency will affect the availability of covariate measurement, and the opportunity to be treated. This chapter therefore aims to descriptively examine some key aspects of the data that may provide insight into whether these issues may be present, and more broadly gain an overview of any other limitations of the data, for example, the extent of missing data and length of available follow up.

6.1.1 Treatment patterns

With the intention of looking at first line treatments, the most obvious comparators for metformin use would be no treatment (assumed to be diet only) or use of sulfonylureas. To feasibly answer a question that makes comparisons between metformin, sulfonylureas and diet only, it is important to know that:

a) All treatments are observed in the data.
b) Treatment decisions are not completely or strongly determined by variables that are important risk factors (key time-dependent confounders) for any of the outcomes of interest, namely cancer, cardiovascular disease and all-cause mortality. I.e. that there are no clear violations of the positivity assumption that may prevent the ability to adjust for time-dependent confounding via IPTW.

To this aim, the observed treatment pathways of patients with newly diagnosed T2DM (as obtained in 5.2) were investigated. Patterns for further intensification were examined to gain an idea of the expected follow up length for the three main treatments of interest.

6.1.2 Frequency of measures of disease severity

As described in 4.4.2, the model fitting process involves separating follow up time into discrete time intervals. It is reasonable to assume that for assessing the presence of particular comorbidities, the absence of a diagnosis code up to a particular interval suggests that the patient does not have that comorbidity. This is because it is likely that a patient would visit the GP either during or shortly after the time at which such a comorbidity develops, to either aid diagnosis or to seek referral to secondary care. Similarly, the absence of a prescription in a particular interval suggests a treatment was not prescribed in that interval, and by making assumptions of continual exposure, or defining exposure to other medications as "use in the last x months", we can obtain non-missing measures of exposure to all types ofGP prescribed medication for each interval. However, for key time varying measures such as BMI and HbA1c there may be intervals in which these measures are not updated, which could be problematic, particularly if there are large differences in frequency of measurement between those on and off treatment. Further, differences in frequency of measures between those initiating and not initiating treatment may be an indication that those not being treated are not attending, and therefore they cannot possibly be treated. Before beginning the modelling process, the availability and frequency of the longitudinal data in the CPRD for the two key time updated covariates of HbA1c and BMI were examined, to establish the presence of any differences in availability of measures, and more generally in attendance, between those initiating and not initiating treatment.

6.2 METHODS

Cleaned files of longitudinal data on BMI, HbA1c and both diabetic and non-diabetic medications were created from those used for diagnosis confirmation (see 5.2.2), to provide longitudinal covariate and exposure information. As well as the existing information about use of metformin and insulin, the group of other OADs were further classified to separately identify sulfonylureas, thiazolidinedione's (TZDs), prandial glucose regulators (glinides), glucagon-like peptide 1 receptor agonists (GLPs) , DPP4 inhibitors (gliptins) and "other" which contained anything else under the BNFcode 6.1.2.3 "Other antidiabetic drugs" .

6.2.1 Treatment patterns

This analysis was restricted to the 98,080 patients for whom follow up started at time of diabetes diagnosis, with no history of cancer prior to diabetes diagnosis. As explained in 5.2.3, their baseline visit (time 0) was defined to be the point at which the diabetes diagnosis was confirmed. The month preceding this time was their "baseline interval". For some patients, treatment initiation may have occurred in this interval, and these patients were considered to be exposed (treated) at baseline. From time 0 onwards, first and second line therapy choices were summarised. Specifically, the absolute number (and percentage) of patients initiating each class of therapy (as monotherapy and specific combinations) at each stage of intensification were calculated. Time between diagnosis and initiation of first treatment, and then time from first treatment to intensification, were also summarised by looking at the mean, median and interquartile range of the follow up time for each choice of treatment.

With specific interest in the treatment options of no treatment, metformin monotherapy, sulfonylurea monotherapy or "other" (which included dual therapy with metformin and sulfonylureas), the distributions of categorical forms of age, BMI, HbA1c, and calendar year of diagnosis were compared between baseline treatments. Missing indicators for BMI and HbA1c were used to investigate the extent of missing data at baseline. Since first line initiation may happen later, mean HbA1c and BMI were also presented per month of follow up, in those initiating and not initiating in each month. This analysis was limited to the first 24 months of follow up and restricted to those with complete data at baseline.

6.2.2 Frequency of measures of disease severity

Because the interest of later chapters is in first line treatment initiation with the most common anti-diabetes agents, for this part of the analysis, patients were only followed while they were on no treatment, metformin monotherapy or sulfonylurea monotherapy. Specifically, follow up time was censored at initiation of any other drug after the first prescription for either metformin or a sulfonylurea.

As explained in 4.4.2.1, when investigating the effect of treatment on outcome using MSMs, only follow up time once patients have complete data on covariates is included. In later chapters, study entry is defined as the start of the interval after complete data are obtained, and the baseline interval (interval 0) is defined to be the interval in which complete data occurs. For patients who have complete data from time of diagnosis, this means their baseline interval

remains, as stated in 6.2.1 and Figure 5.3, to be the month preceding the confirmed diabetes diagnosis. This is explained in more detail in sections 7.2.4 and 7.2.5. To ensure that the frequency of measures was examined within the relevant follow up for the main analyses, we used the new definition of study entry for this part of the analysis, and ended follow up at the earliest of death, transfer from GP practice, intensification of treatment (as described above), or end of data collection.

Looking separately at BMI and HbA1c, the number of measures per person year of follow up were estimated in the overall population. Then, for each patient, the average time between measures during their follow up was calculated. This was done by dividing total person time for each individual by the number of measures taken during their follow up. This was repeated looking in the periods both before and after (if applicable) initiation of first line treatment with either sulfonylureas or metformin.

In those patients who initiated treatment, and therefore had a relevant date as reference, time between the closest measurement of BMI/HbA1c and treatment date was calculated and summarised to assess whether these measures could plausibly represent the value that may have informed the decision to initiate treatment.

Additionally, the number of measures since study entry, and the time since last measure, were examined at specific time points (6, 12 and 24 months after study entry) in both those initiating at that time point (initiators), and in those who were still untreated (non-initiators) at that time point. This was done to better compare frequency and timeliness of measures between treated and untreated individuals.

To further investigate any differences in frequency of measures between initiators and non-initiators, the number of months in the previous calendar year where a patient received at least one consultation was calculated. The distribution of this variable was then examined between initiators and non-initiators at the same time points of 6, 12 and 24 months after study entry, to see whether any differences in frequency of measures could be explained by increased or decreased attendance. Note that because data were available in the CPRD before actual study entry, this could still be estimated for only 6 months into study follow up. Finally, the change in value from baseline was also examined in initiators and non-initiators that had an updated HbA1c or BMI within 3 months of the 6, 12 and 24 month time points.

6.3 RESULTS

6.3.1 Treatment patterns

Mean follow up for the 98,080 patients was 64.6 months (SD 47.8 months, median 55 months, IQR 25-96 months).

Numbers initiating each class of therapy, and summary statistics for time between diagnosis and initiation of first line therapy are displayed in Table 6.1. Metformin monotherapy was the most popular first line therapy choice, with 75% of patients initiating this as their first anti-diabetes agent. However, the time to initiation of metformin varied. In those patients who initiated metformin, the mean number of months from time of diagnosis to initiation of metformin was 12.8 months, with a median time of 2 months, IQR 1-16 months. The second most popular choice was a sulfonylurea, though if this was the first line therapy used, the time that it was started tended to be much closer to diagnosis (mean 7.8 months, median 1 month, IQR 1-4 months).

Options for second line intensification and time between first and second therapy are also summarised in Table 6.1. Sulfonylureas were the most common therapy choice for second line treatment, with 55% of all those who intensified moving onto them (70% if the first line therapy was metformin). The median time between first treatment and intensification in those who intensified was 20 months (IQR of 7-41 months). In those who didn't start a second therapy, median follow up time after first line therapy initiation was 31 months (IQR 13-58 months). The use of newer agents such as DPP4 inhibitors appeared low, however this is likely to be a reflection of length of time since such medications were approved. Restricted to 2005 onwards, metformin became an even more popular first line therapy choice with around 80% of patients initiating on this drug. Also, use of DPP4 inhibitors for treatment intensification after metformin increased. See appendix 10 for more detailed post 2005 results.

In the overall study population, 14,388 patients went onto a third, and 4820 to a fourth line treatment. Only 1,223 patients initiated further therapies (fifth line and beyond) during their follow up. The most common choices for these third and fourth line therapies were insulin, DDP4 inhibitors and GLPs (for 4^{th} line).

When looking at different first line treatments by levels of different covariates, those who were older at time of diagnosis were more likely to remain untreated, or be treated with a sulfonylurea rather than metformin. As would be anticipated, patients being treated with medication straight away had higher HbA1c at time of diabetes diagnosis, with the highest

Hba1c in those initiating a sulfonylurea. Patients initiating sulfonylureas also had slightly lower BMI at the time of diagnosis than those initiating metformin (Table 6.2–Table 6.4). Overall, 41% of patients had missing data on HbA1c at baseline, and 23% had missing data on BMI. The finding of metformin and sulfonylurea initiators having higher mean HbA1c, and sulfonylurea initiators having lower mean BMI remained when this was examined at each month for the first 2 years of follow up (Table 6.5, Table 6.6).

	First line treatment choice		Followed by intensification with …					
			All those who intensify		if first therapy was metformin		if first therapy was a sulfonylurea	
	First therapy N (%)	Average time from diagnosis to start of therapy (months) Mean, Median, IQR	Second therapy N (%)	Average time from first to second therapy (months) Mean, Median, IQR	Second therapy N (%)	Average time from first to second therapy (months) Mean, Median, IQR	Second therapy N (%)	Average time from first to second therapy (months) Mean, Median, IQR
Metformin	53134 (73.8)	12.8 , 2 , 1 – 16	7310 (21.1)	27.9 , 19 , 7 – 40			6,880 (86.3)	28.1 , 20 , 7 – 40
Sulfonylurea	12436 (17.3)	7.8 , 1 , 1 – 4	16975 (49.0)	25.5 , 17 , 6 – 37	16754 (69.5)	25.6 , 17 , 6 – 37		
Insulin	434 (0.6)	8.3 , 1 , 1 – 6	1458 (4.2)	25.7 , 15 , 4 – 38	360 (1.5)	24.2 , 17 , 5 – 37	598 (7.5)	23.5 , 12 , 3 – 35
TZDs	160 (0.2)	17.4 , 3 , 1 – 30.5	3389 (9.8)	34 , 27 , 11 – 51	2727 (11.3)	33.0 , 26 , 10 – 50	192 (2.4)	38.4 , 29 , 11 – 55
Glinides	86 (0.1)	10.3 , 1 , 1 – 12	284 (0.8)	18 , 11 , 4 – 25	199 (0.8)	17.9 , 11 , 4 – 25	46 (0.6)	18.5 , 12 , 4 – 29
GLPs	9 (0.01)	32.2 , 16 , 5 – 53	399 (1.2)	32.4 , 24 , 10 – 48	268 (1.1)	31.1 , 24 , 10 – 46	1 (0.0)	9 , 9 , 9 – 9
DPP4	87 (0.1)	18.4 , 3 , 1 – 28	3704 (10.7)	34.5 , 27 , 11 – 50	3001 (12.4)	33.6 , 26 , 11 – 49	97 (1.2)	42.0 , 31 , 15 – 60
Other	100 (0.1)	8.7 , 1 , 1 – 7	794 (2.3)	32.0 , 25 , 9 – 46	592 (2.5)	31.3 , 24 , 9 – 44	83 (1.0)	28.3 , 20 , 8 – 41
Metformin/Sulf combination	4567 (6.3)	3.4 , 1 , 1 – 1	11 (0.0)	9.0 , 4 , 2 – 11				
Metformin/Insulin	159 (0.2)	2.7 , 1 , 1 – 1	22 (0.1)	19.9 , 4 , 1 – 14			22 (0.3)	19.9 , 4 , 1 – 14
Metformin/Other	344 (0.5)	5.8 , 1 , 1 – 1	53 (0.2)	30.6 , 18 , 7 – 46			50 (0.6)	32.3 , 20 , 8 – 47
Sulfonylurea/Insulin	220 (0.3)	3.9 , 1 , 1 – 1	50 (0.1)	15.2 , 8 , 1 – 17	50 (0.2)	15.2 , 8 , 1 – 17		
Sulfonylurea/Other	45 (0.06)	10 , 1 , 1 – 1	143 (0.4)	27.9 , 21 , 5 – 43	138 (0.6)	27.0 , 20 , 4 – 42		
Insulin/Other	3 (0.0)	7 , 1 , 1 – 19	24 (0.1)	28.8 , 26 , 3 – 39	5 (0.0)	23.8 , 10 , 1 – 31	4 (0.1)	43.3 , 31 , 3 – 84
Other dual combination	0 (0.0)		19 (0.1)	22.7 , 18 , 3 – 33	12 (0.0)	24.3 , 18 , 4 – 33	0 (0.0)	
3 or more	222 (0.3)	3 , 1 , 1 – 1	5 (0.0)	31.6 , 15 , 14 – 51	2 (0.0)	14.5 , 15 , 14 – 15	2 (0.0)	32.5 , 33 , 14 – 51
Sub Total	**72,006**	**11.2 , 1 , 1 – 12**	**34,640**	**28.0 , 20 , 7 – 41**	**24,108**	**27.5 , 20 , 7 – 41**	**7,975**	**28.2 , 19 , 6 – 40**
None	26,074	43.7, 34,12-64	37,366	39.8 , 31 , 13 – 58	29,026	39.9 , 30 , 13 - 57	4,461	43.3 , 30 , 12-63
Total	**98,080**		**72,006**		**53,134**		**12,436**	

Table 6.1 Summary of first and second line treatment options for patients diagnosed with T2DM.

Time in months for "none" represents time until end of follow up.

Medication at diagnosis	Age (years)				
	30-49	**50-59**	**60-69**	**70+**	**Total**
None	9,173(49.3)	13,765(56.8)	17294(62.3)	18076(65.8)	58,317(59.5)
Metformin monotherapy	6,561(35.3)	7364(30.4)	6995(25.2)	5163(18.8)	26,083(26.6)
Sulfonylurea monotherapy	1,276(6.9)	1692(7.0)	2203(7.9)	3088(11.2)	8,259(8.4)
Other	1,601 (8.6)	1415 (5.8)	1263 (4.6)	1151 (4.2)	5,430 (5.5)
Total	18,611	24,236	27,755	27,478	98,080

Table 6.2 N (%) of patients initiating each kind of therapy at time of diabetes diagnosis, presented by categories of age at time of diagnosis.*

* within the one month interval ("baseline interval") in which diagnosis of diabetes is confirmed.

Medication at diagnosis	BMI (kg/m²)					
	<25	**25-29**	**30-34**	**35+**	**missing**	**Total**
None	4,770(54.7)	14,701(62.1)	14,247(62.9)	11,601(58.8)	12,989(55.7)	58,308(59.4)
Metformin monotherapy	1,635(18.8)	6,026(25.4)	6,743(29.8)	6,998(35.5)	4,681(20.1)	26,083(26.6)
Sulfonylurea monotherapy	1,633(18.7)	1,751(7.4)	713(3.1)	345(1.7)	3,817(16.4)	8,259(8.4)
Other	676 (7.8)	1,214 (5.1)	937 (4.1)	786 (4.0)	1,817 (7.8)	5,430 (5.5)
Total	8,714	23,692	22,640	19,730	23,304	98,080

Table 6.3 N (%) of patients initiating each kind of therapy at time of diabetes diagnosis, presented by categories of BMI at time of diagnosis.*

* within the one month interval ("baseline interval") in which diagnosis of diabetes is confirmed.

Medication at diagnosis	HbA1c (%)							
	<6%	**6-6.5%**	**6.5-7%**	**7-8%**	**8-9%**	**>=10%**	**missing**	**Total**
None	3,148 (89.5)	5,354 (86.3)	7,891 (82.7)	7,848 (67)	4,574 (40.5)	2,682 (18.6)	26,811 (64.8)	58,308 (59.4)
Metformin monotherapy	281 (8.0)	724 (11.7)	1,493 (15.6)	3,355 (28.6)	5,313 (47.1)	7,530 (52.2)	7,387 (17.9)	26,083 (26.6)
Sulfonylurea monotherapy	68 (1.9)	83 (1.3)	98 (1.0)	330 (2.8)	807 (7.1)	1,825 (12.6)	5,048 (12.2)	8,259 (8.4)
Other	22 (0.6)	44 (0.7)	62 (0.6)	189 (1.6)	595 (5.3)	2,396 (16.6)	2,122 (5.1)	5,430 (5.5)
Total	3,519	6,205	9,544	11,722	11,289	14,433	41,368	98,080

Table 6.4 N (%) of patients initiating each kind of therapy at time of diabetes diagnosis, presented by categories of HbA1c at time of diagnosis.*

* within the one month interval ("baseline interval") in which diagnosis of diabetes is confirmed.

Months since diagnosis	Non initiator						Metformin monotherapy						Sulfonylurea monotherapy					
	N	Mean	SD	50%	25%	75%	N	Mean	SD	50%	25%	75%	N	Mean	SD	50%	25%	75%
1	31,497	7.37	1.59	6.9	6.4	7.8	18,696	9.48	2.26	9.2	7.6	11.1	3,211	10.44	2.49	10.4	8.5	12.2
2	31,704	7.23	1.47	6.8	6.4	7.6	1,484	9.09	1.92	8.8	7.6	10.4	256	9.90	2.22	9.7	8.1	11.3
3	32,900	7.06	1.33	6.7	6.3	7.4	1,587	8.69	1.73	8.3	7.4	9.6	224	9.18	1.83	9.1	7.8	10.4
4	34,788	6.85	1.16	6.6	6.2	7.2	1,594	8.42	1.52	8.0	7.4	9.2	177	8.77	1.83	8.5	7.6	9.7
5	35,285	6.72	1.04	6.6	6.1	7.1	1,019	8.32	1.47	8.0	7.3	9.1	122	8.76	1.57	8.45	7.6	9.5
6	35,142	6.65	0.97	6.5	6.1	7.0	707	8.10	1.43	7.8	7.2	8.6	99	8.39	1.73	8.0	7.3	9.3
7	34,718	6.58	0.91	6.5	6.0	6.9	663	7.94	1.34	7.7	7.1	8.5	94	8.14	1.63	7.8	7.2	8.7
8	34,173	6.54	0.88	6.5	6.0	6.9	517	8.08	1.41	7.8	7.2	8.7	84	8.15	1.66	7.8	7.1	8.7
9	33,599	6.51	0.86	6.4	6.0	6.9	481	7.88	1.32	7.6	7.1	8.4	44	7.94	1.51	7.9	7.15	8.55
10	32,897	6.49	0.84	6.4	6.0	6.8	469	7.94	1.44	7.6	7.1	8.4	68	8.09	1.74	7.7	7.1	8.85
11	32,216	6.47	0.83	6.4	6.0	6.8	462	7.89	1.43	7.6	7.0	8.3	49	7.99	1.57	7.8	7.0	8.6
12	31,435	6.45	0.82	6.4	6.0	6.8	547	7.86	1.32	7.5	7.0	8.3	51	7.97	1.62	7.6	7.0	8.6
13	30,728	6.43	0.80	6.4	5.9	6.8	547	7.87	1.28	7.6	7.0	8.4	58	7.74	1.48	7.5	6.9	8.5
14	30,035	6.42	0.80	6.4	5.9	6.8	462	7.90	1.29	7.6	7.1	8.3	43	8.42	1.91	8.2	7.1	9.0
15	29,363	6.41	0.79	6.3	5.9	6.8	412	7.81	1.21	7.6	7.0	8.3	56	7.97	1.44	7.7	7.2	8.5
16	28,666	6.41	0.79	6.3	5.9	6.8	400	7.85	1.32	7.6	7.0	8.3	31	7.95	1.62	7.9	6.7	9.2
17	28,019	6.40	0.78	6.3	5.9	6.8	391	7.85	1.29	7.5	7.0	8.4	39	7.59	1.16	7.6	6.6	8.1
18	27,325	6.40	0.78	6.3	5.9	6.8	394	7.81	1.24	7.6	7.1	8.3	36	7.83	1.68	7.6	6.7	8.6
19	26,715	6.39	0.77	6.3	5.9	6.8	377	7.81	1.31	7.6	7.1	8.2	53	7.77	1.67	7.5	6.7	8.1
20	26,105	6.39	0.78	6.3	5.9	6.8	314	7.84	1.22	7.6	7.0	8.3	46	7.60	1.44	7.4	7.0	8.3
21	25,501	6.39	0.77	6.3	5.9	6.8	348	7.85	1.44	7.6	6.9	8.3	43	7.87	1.66	7.7	6.5	8.7
22	24,941	6.39	0.77	6.3	5.9	6.8	297	7.90	1.46	7.6	7.0	8.4	35	8.01	1.72	7.7	7.07	8.4
23	24,344	6.39	0.77	6.3	5.9	6.8	326	7.83	1.54	7.5	6.9	8.3	45	7.64	1.28	7.5	6.9	8.2
24	23,739	6.39	0.77	6.3	5.9	6.8	319	7.87	1.31	7.6	7.1	8.5	37	8.11	1.59	7.9	6.9	9.0

Table 6.5 Distribution of HbA1c (%) in those initiating metformin and sulfonylureas vs non initiators in the first 24 months of follow up after diabetes diagnosis.

Months since diagnosis	Non initiators						Metformin monotherapy						Sulfonylurea monotherapy					
	N	Mean	SD	50%	25%	75%	N	Mean	SD	50%	25%	75%	N	Mean	SD	50%	25%	75%
1	45319	31.8	6.2	30.9	27.6	35.1	21402	33.1	6.8	32.0	28.4	36.7	4442	27.2	5.3	26.4	23.7	29.8
2	45757	31.7	6.2	30.8	27.5	35.0	1847	32.7	6.4	31.9	28.1	36.1	392	27.2	4.9	26.1	24.0	29.5
3	44804	31.6	6.2	30.7	27.4	34.8	2015	32.5	6.1	31.4	28.3	35.9	316	27.1	4.8	26.6	23.9	29.3
4	43409	31.4	6.1	30.5	27.3	34.6	1794	33.0	6.5	31.7	28.4	36.3	264	27.2	5.1	26.6	24.0	29.5
5	42458	31.2	6.1	30.4	27.1	34.4	1137	32.6	6.1	31.6	28.1	36.0	164	27.8	5.3	26.7	24.1	30.8
6	41607	31.1	6.0	30.2	27.0	34.3	765	32.8	6.3	32.0	28.1	36.3	138	28.1	4.8	27.5	24.8	29.7
7	40703	31.0	6.0	30.1	26.9	34.2	733	32.7	6.4	31.7	28.5	36.3	117	28.2	5.1	27.3	25.2	30.3
8	39837	30.9	6.0	30.1	26.8	34.1	560	33.2	6.1	32.3	28.9	36.4	115	27.7	5.7	26.7	23.8	30.8
9	38972	30.8	6.0	30.0	26.8	34.0	530	33.4	6.5	32.4	29.2	36.8	64	27.7	4.6	26.9	24.5	30.3
10	38050	30.7	6.0	29.9	26.7	33.9	508	33.1	6.3	32.1	28.7	36.7	98	28.5	5.3	27.0	24.7	31.2
11	37182	30.7	6.0	29.8	26.6	33.9	506	33.0	6.1	32.0	29.1	36.4	62	27.8	3.5	27.8	25.3	30.2
12	36147	30.6	6.0	29.8	26.6	33.8	602	32.8	6.0	32.2	28.4	36.2	72	27.7	4.6	27.3	24.0	29.9
13	35150	30.6	6.0	29.8	26.6	33.7	591	32.9	6.0	32.3	28.7	36.2	72	28.2	5.0	27.3	24.3	31.8
14	34224	30.5	6.0	29.7	26.5	33.7	510	33.0	6.3	32.1	28.7	35.9	63	28.0	4.6	27.4	24.8	30.0
15	33335	30.5	6.0	29.7	26.5	33.7	465	32.6	6.6	31.4	28.2	35.9	76	29.2	4.9	28.9	25.3	32.6
16	32479	30.5	6.0	29.7	26.4	33.6	443	32.7	6.1	31.8	28.5	36.1	46	28.7	5.0	28.4	25.5	31.3
17	31657	30.4	6.0	29.7	26.4	33.6	436	33.2	6.3	32.0	28.7	36.8	53	28.2	5.4	27.6	24.1	31.1
18	30817	30.4	5.9	29.6	26.4	33.6	426	33.1	6.6	31.8	28.7	36.2	53	27.5	4.8	26.9	24.2	30.8
19	30074	30.4	5.9	29.6	26.4	33.5	403	32.8	6.1	31.7	28.6	36.2	66	28.6	4.9	27.3	25.3	31.2
20	29348	30.4	5.9	29.6	26.4	33.6	336	32.8	6.5	31.4	28.3	36.3	54	27.8	4.8	27.8	25.0	29.6
21	28641	30.4	5.9	29.6	26.3	33.5	375	32.7	6.4	31.5	28.4	35.9	52	27.7	4.1	27.3	25.1	30.2
22	27976	30.3	5.9	29.6	26.3	33.5	323	32.9	6.3	31.9	28.7	36.1	47	27.6	5.2	27.1	23.3	30.2
23	27285	30.3	5.9	29.5	26.3	33.5	350	32.5	5.6	32.2	28.5	35.5	56	28.7	6.1	26.8	24.8	32.8
24	26567	30.3	5.9	29.5	26.3	33.4	344	33.0	6.3	32.4	28.3	36.6	50	27.9	4.7	27.3	25.2	30.8

Table 6.6 Distribution of BMI (kg/m^2) in those initiating metformin and sulfonylureas vs non initiators in the first 24 months of follow up after diabetes diagnosis.

6.3.2 Visit frequency

Redefining study entry to be the time after complete data were obtained (as explained in section 6.2.2) resulted in 59,079 patients contributing to this analysis. 7,389 of these patients were treated with metformin or a sulfonylurea from baseline. 23,428 did not receive prescriptions for metformin or a sulfonylurea at any time during their follow up and were assumed to be managing their diabetes with diet alone.

Excluding the baseline measures, the rate of measures of HbA1c over all of follow up was 1.60 per person year of follow up, and 1.61 per person year for BMI. This rate translates to an mean time between measures of 7.49 months. Alternatively, calculating the rate of measurements in each individual and converting this to an average time between measurements, gives a mean time for HbA1c of 7.6 months, with a median of 6.5 months, IQR (5–9 months). Restricting this to the time before any treatment occurs, and ignoring those who were treated or lost to follow up in the same interval in which their first measure occurred, the mean time between measures pre-treatment was 6.8 months, median 6 months, IQR 3-9 months (Table 6.7). For BMI, the equivalent numbers were very similar (Table 6.8).

Looking only in those who received treatment, 2,669/35,651 had no further HbA1c measures after treatment initiation, and 503/36,561 initiated therapy in their last month of follow up. However in the remaining 32,479, the mean time between measures of HbA1c after treatment was initiated was 8.14 months (Table 6.7). A longer gap between measures after treatment initiation was also observed for BMI (Table 6.8).

Time between intervals with updated HbA1c	N	Mean (months)	Std. Dev. (Months)	Median	IQR
Overall	59,079	7.62	5.8	6.5	4.9 - 8.8
Pre- treatment*	48,481	6.86	6.4	5.9	3.3 – 8.5
Post – treatment**	32,479	8.14	5.0	7.0	5.6 - 9.2

Table 6.7 Summary of number of months between HbA1c records in CPRD data after study entry, presented by overall, pre and post treatment periods.

*excludes 3,209 who have a time between intervals of 0 as these are people who are either treated or lost to follow up in the same interval in which they have their first measure.
**excludes 2,669 patients with no further HbA1c measures after treatment initiation and 503 that initiate in their last month.

Time between intervals with updated BMI	N	Mean (months)	Std. Dev. (Months)	Median	IQR
Overall	59,079	7.87	5.8	6.8	4.8 - 9.5
Pre- treatment*	48,481	6.70	5.9	5.6	3.2 - 8.5
Post – treatment	31,348	9.69	6.7	8.0	5.9 - 11.7

Table 6.8 Summary of number of months between BMI records in CPRD data after study entry, presented by overall, pre and post treatment periods.

*excludes 3,209 who have a time between intervals of 0 as these are people who are either treated or lost to follow up in the same interval in which they have their first measure.
**excludes 3800 patients with no further BMI measures after treatment initiation and 503 that initiate in their last month.

This suggests that, on average, once patients had been diagnosed with T2DM and have had at least one measure taken (at study entry), they are re-measured just over every six months, with the possibility that once treated, the monitoring gets slightly less frequent.

The number of measurements taken between baseline and 6, 12 and 24 month intervals were summarised separately for those initiating and not initiating in those intervals (Table 6.9). For both HbA1c and BMI, those initiating in a given interval had more measures on average recorded up to that point than those not treated.

	Non initiator [Mean (SD) Median (IQR)]			**Initiator [Mean (SD) Median (IQR)]**		
	6 months	**12 months**	**24 months**	**6 months**	**12 months**	**24 months**
HbA1c	2.15 (0.91) 2 (1-3)	2.79 (1.12) 3 (2-4)	3.97 (1.50) 4 (3-5)	2.77 (0.90) 3 (2-3)	3.38 (1.15) 3 (3-4)	4.81 (1.54) 5 (4-6)
BMI	2.84 (1.80) 3 (2-3)	3.60 (2.08) 3 (2-4)	5.02 (2.56) 5 (3-6)	2.97 (1.35) 3 (2-4)	3.99 (2.22) 4 (3-5)	5.36 (2.09) 5 (4-7)

Table 6.9 Mean, SD, median and IQR of the total number of measures of HbA1c/BMI recorded by 6, 12 and 24 months after study entry, in those initiating and not initiating treatment at those time points.

To understand how these measures were distributed in terms of time to treatment, the time between treatment and the closest updated measurement were summarised. For both HbA1c and BMI, the majority of patients who were treated had an updated measure within three months prior to the treatment date (Table 6.10, Table 6.11). However, in those who were not

treated, the mean time in months since the last measure was much greater, and the longer they remained untreated, the longer the time since the last measure appeared to be (Table 6.11).

Timing of closest measurement entry to treatment initiation	HbA1c N (%)	BMI N (%)
< 3 months	32,301 (90)	28,314 (79)
3-6 months	1,977 (6)	3,479 (10)
6-12 months	1,039 (3)	2,707 (8)
> **12 months**	499 (1)	1,316 (4)
Total	35816	35816

Table 6.10 Timing of closest HbA1c and BMI to treatment initiation. Includes data from all patients who are started on treatment at any time during follow up.

	Non initiator [Mean (SD) Median (IQR)]			Initiator [Mean (SD) Median (IQR)]		
	6 months	**12 months**	**24 months**	**6 months**	**12 months**	**24 months**
HbA1c	3.93 (2.78) 4 (2-6)	5.30 (4.60) 4 (1-9)	6.73 (6.75) 5 (2-10)	1.47 (1.97) 1 (0-2)	1.58 (2.78) 1 (0-1)	1.67 (3.13) 1 (0-1)
BMI	4.19 (3.96) 5 (2-6)	5.62 (5.30) 5 (1-10)	6.65 (6.68) 5 (2-10)	2.37 (3.26) 1 (0-5)	2.58 (4.14) 0 (0-4)	2.54 (4.55) 0 (0-3)

Table 6.11 Time in months since last measure, at 6, 12 and 24 months after study entry, separately by those initiating and not initiating treatment at those times.

Despite this finding, for all three time points examined, the number of consultations in the previous year (simplified to maximum of 1 visit per month) was not appreciably different between initiators and non-initiators (Table 6.12). The changes (as a percentage of the baseline value) in HbA1c from baseline to each time interval examined were larger in those initiating treatment than those not (Table 6.13). For BMI, both the mean and median percentage change from baseline was approximately 1% larger in patients not initiating treatment at all three intervals, though overall the percent change was relatively small. For example, at a starting BMI

of 30, after 24 months the average change was only 4% in initiators, and 5% in non-initiators, resulting in an absolute change in BMI of 1.2 in initiators and 1.5 in non-initiators.

	Non initiator [Mean (SD) Median (IQR)]			Initiator [Mean (SD) Median (IQR)]		
	6 months	**12 months**	**24 months**	**6 months**	**12 months**	**24 months**
Number of months in previous year that at least one consultation occured	8.7 (2.6) 9 (7-11)	8.0 (2.8) 8 (6-10)	8.9 (2.3) 9 (7-11)	9.2 (2.4) 9 (8-11)	8.2 (2.7) 8 (6-10)	8.7 (2.6) 9 (7-11)

Table 6.12 Mean, SD median and IQR of the number of months in the previous year in which at least one consultation occurred, recorded at 6 12 and 24 months after study entry, in those initiating and not initiating at those time points.*

*because data source has information on consultations for at least 1 year prior to study entry for all patients, this is estimable for 6 months into follow up.

	Non initiator [Mean (SD) Median (IQR)]			Initiator [Mean (SD) Median (IQR)]		
Absolute % change from baseline in:	**6 months**	**12 months**	**24 months**	**6 months**	**12 months**	**24 months**
HbA1c (%)	8.6 (9.0) 5.7 (3.0-10.8)	8.6 (9.3) 5.8 (2.9-11.1)	9.4 (10.0) 6.3 (3.1-12.3)	13.1 (14.9) 8.6 (4.1-16.7)	15.5 (16.5) 11 (5.4-20.4)	21.4 (20.3) 16.2 (7.0-28.4)
BMI	4.1 (3.77) 3.2 (1.5-5.7)	4.8 (4.5) 3.5 (1.6-6.5)	5.1 (5.0) 3.8 (1.7-6.9)	3.1 (2.8) 2.4 (1.1-4.1)	3.8 (5.0) 2.9 (1.3-4.5)	4.0 (3.4) 3.3 (1.6-5.3)

Table 6.13 Mean, SD, median and IQR of the absolute percentage change between study entry and most recent HbA1c/BMI recorded by 6, 12 and 24 months after study entry, in those initiating and not initiating treatment at those time points.*

*must have occurred within 3 months of the time point for HbA1c, and 4 months for BMI to ensure mean change is not reduced by patients who have not had any updated measurement.

6.4 DISCUSSION

The results confirmed that, as per UK guidelines, metformin is the preferred first line treatment out of all options. In addition, the changes in prescribing over the last 20 years in terms of a move towards metformin over sulfonylureas, and the introduction of new 2nd line therapies in the last decade (e.g. DPP4) were apparent when comparing between categories of calendar year or restricting to data from post 2005.

These findings are consistent with a recently published study by Sharma et al. [16] which looked at incidence and prevalence of diabetes and the patterns of pharmacological treatment between 2000 and 2013 in a different UK primary care database (The Health Information Network (THIN) database). The study found that 63% of newly diagnosed patients initiated therapy at some point in the follow up, compared to 73% in our study. Though, when restricting our study to post 2005 only, this number was 68%. The authors' findings in the period 2005 to 2013, showed an increase in metformin prescribing as first line therapy from around 80% to 90%, and a decrease in sulfonylurea usage from around 15% to 6%. These results are highly consistent with our findings over a similar period (appendix 10). Their findings on second line treatments are also very similar.

In those who do initiate metformin, the median time to initiation was 2 months. However, there were a substantial number of people who started later, or did not start at all. There is the possibility that patients appearing to not be treated are getting pharmacological therapy elsewhere, although for diabetes the majority of prescribing is done in primary care so the risk of misclassification of treatment is relatively small [12].

Overall, the findings of this analysis provide support for the hypothesis that not all patients with diagnoses of T2DM initiate medication immediately, and that there is some variation in choice of first line diabetes therapy. Therefore it should be possible to model time-varying treatment and compare diet only, metformin monotherapy and sulfonylurea monotherapy.

Age, HbA1c and BMI all appeared to be strong predictors of treatment, even when relatively simply categorised. However, there did not appear to be a group that were never or always treated, suggesting that there should be no outright positivity violations. Having said this, it is acknowledged that only a small selection of possible confounders were examined and only in univariable analyses. The prescribing of glucose lowering medications in CPRD appeared to reflect guidelines [13], but suggested that there was enough overlap in the distributions of potential confounders for those starting and not starting treatment, that MSM with IPTW may be viable to implement. It is possible that there will be some near positivity violations not identified here. For example, there may be issues at the extremes of the continuous distributions of these or other potential confounding variables, or when multivariable associations with treatment are examined. This will be considered in the modelling process as previously explained in relation to the bias-variance trade off (see sections 4.4.2.3-4.4.2.4).

The above analysis also suggested that there were a substantial amount of data missing for key time-dependent confounders at time of diabetes diagnosis. As explained in 4.4.2.1 and 6.2.2 ,

the approach taken will be define study entry to the closest time after diagnosis that complete data are available, and use the method of last one carried forward (LOCF) to impute values going forward in time. How well the inverse probability of treatment weighting will account for post baseline confounding when using LOCF will depend on how often the measures are updated, since we wish to capture the values that might influence the risk of outcome at the time the treatment decisions are made. The second set of analyses in this chapter aimed to describe the frequency of measurements for HbA1c and BMI, along with general measures of visit frequency, to gain an idea of whether there are enough data on such measures to obtain valid weights.

Overall, both HbA1c and BMI, once measured, had a mean time between measurements of just under 8 months, with a median of nearer 6 months. In those eventually treated, the time between measures was slightly shorter before treatment than after, though this was not formally tested, and the difference was small. This could suggest that HbA1c and BMI are more frequently monitored during the time when a patient is attempting to control their diabetes with a diet/lifestyle intervention, though it could also be a reflection of the fact that those more closely monitored are more likely to be treated. This potential imbalance was accounted for by looking at the number of measures of HbA1c and BMI taken by 6, 12 and 24 months after study entry in those initiating or not initiating treatment in those months. Here, there was a trend towards a higher number of measures in those initiating treatment at the given time points than those not initiating treatment, though the differences were relatively small.

HbA1c is a measure that is representative of glucose control over approximately 3 months, and changes in BMI are likely to take a longer to occur. Within those treated, the majority of patients (90% for HbA1c and 80% for BMI) had a measure recorded in the 3 months prior to treatment initiation. This suggests that, ignoring issues of measurement error, the value that would be used to estimate probability of being treated is likely representative of the value used by the GP when making the decision to initiate pharmacological therapy. In those not treated, we are limited to making this comparison at specific time points. In doing this, it was apparent that the closest measures for both HbA1c and BMI were further in the past in those not initiating than initiating at 6, 12 and 24 months after study entry. If this is because they are not visiting the GP, then the patient can neither have a measurement or be treated. This could be very problematic for two main reasons. Firstly, using LOCF as a method to estimate the probability of treatment in each interval is likely to result in an estimated weight that does not correctly adjust for the time-dependent confounding, since the risk of the outcome at each time interval may not be correctly represented in the patients not initiating treatment. Secondly, if the reason for non-attendance is related to risk of outcome, then additional bias could be introduced, since patients not being

treated are at a systematically different risk of the outcome than those who are treated. Having said this, for BMI, the mean time since the last measure was not so large that it was likely have changed radically. For example, in those not initiating by 2 years after diabetes diagnosis, the mean time since the last BMI was 6.65 months. A large change in BMI over this interval is unlikely to occur, except perhaps following bariatric surgery.

The finding of less measures and longer time since measures in those not initiating treatment may also be reflective of the GP only recording the measure if it has substantially changed from the previous recorded measure. If a patient is visibly stable while controlling their diabetes with diet alone then they may not have as many updated measures. If this is the case, then using LOCF may still provide measures that are broadly representative of the value at the time the decision to remain untreated is made, and as such, correctly represent risk of outcome at this time. Smaller percentage changes in HbA1c from baseline were observed at 6, 12 and 24 months from study entry in patients not initiating treatment compared to those initiating treatment, when restricted to patients that had an updated measure within 3 months of the given time point. For BMI, the percentage change was marginally bigger in patients not being treated, but only by about 1% at each time point. These observations are consistent with the assumption that those not initiating treatment have more stable covariates, so gives some reassurance that LOCF may be a reasonable approach. At all three time points of 6, 12 and 24 months after study entry, the number of months in the previous year in which at least one consultation occurred was similar between initiators and non-initiators. This provides further assurance that the patients not initiating treatment were still visiting the GP, and that they had similar opportunity to be treated as patients that did initiate treatment. Having said this, these analyses cannot remove all concern that the issue of differing visit frequency may impact the results obtained in subsequent chapters.

It should also be noted that this analysis did not consider the limitations that may be encountered due to completly missing data either for HbA1c and BMI, or other time invariant (for the purposes of this research) variables such as smoking and alcohol intake. This will be discussed in more detail in the next chapter.

6.5 CHAPTER SUMMARY

This chapter has provided an overview of diabetes treatment patterns in this base cohort, and frequency of key time updated variables have been examined descriptively to gain initial insight

into whether use of causal methodology, in particular the implementation of MSMs with IPTW will be feasible.

The key findings from the work presented in this chapter are as follows:

1. The most common first line treatment for type 2 diabetes was metformin, though not all patients who do initiate metformin initiate at time of diagnosis. Therefore, in order to compare no treatment to treatment with metformin, time-dependent confounders affected by prior treatment should be considered.

2. The total mean follow up in those never treated was 43 months. In patients who ever initiated a first line therapy, the mean time to first therapy was 11 months. Second line treatment or end of follow up occurred on average 30 months after initiation of first line therapy. This suggests there should be sufficient follow up to capture many outcomes if we censor at treatment intensification and apply inverse probability of censoring weights to adjust for informative loss to follow up. However, sensitivity analyses will be necessary to look at treatment effects in patients with longer follow up or exposure time to investigate long term outcomes such as cancer in more detail.

3. Treatment decisions appear strongly associated with HbA1c levels as expected, and to a lesser extent BMI. Assuming HbA1c and BMI are also associated with the outcomes of interest in the later chapters, these two key time-dependent confounders will need to be adjusted for. The strength of the associations with treatment, particularly for HbA1c, is likely to cause issues with large weights, and therefore careful consideration of how to include these variables in the model to balance sufficient control for the confounding with minimising the impact of positivity violations will be needed.

5. Comparison of consultation rates suggested that patients not initiating treatment visit their GP at roughly the same frequency as patients who initiate treatment. The frequency and timing of both HbA1c and BMI appears to be sufficient to calculate the IPTW (and IPCW). In those who are treated, the measures that will contribute to estimating the probability of treatment appear to be measured close enough to the time of treatment to be broadly representative of the values influencing the risk of outcome at the time of the treatment decision. It is less clear whether the same is true for those who remain untreated at a particular time, and the potential impact of this on the results of subsequent analysis will be considered.

7 METFORMIN AND RISK OF CANCER: AN APPLICATION OF MARGINAL STRUCTURAL MODELS WITH INVERSE PROBABILITY OF TREATMENT WEIGHTING

7.1 INTRODUCTION, AIMS AND OBJECTIVES

As established by the systematic review in chapter 2, many studies have previously examined the association between metformin use and risk of cancer. However, most were at risk of various biases, and none to date have expressly addressed the potential issue of time-dependent confounding affected by prior treatment when time varying treatment is modelled.

In order to examine the true causal association between use of metformin and cancer incidence in patients with T2DM, an ideal randomised controlled trial might randomise patients with newly diagnosed T2DM to receive either metformin monotherapy or placebo. To ensure everyone receives standard care, both groups could be additionally advised to follow a diet and exercise regime to further control their diabetes. Providing that there is then sufficient follow up to capture development of cancer, and patients adhere to their randomised treatment, this trial would, as best is possible, remove any systematic differences between those using and not using metformin prior to any further treatment initiation. However, such a trial is not feasible, since it is unethical to withhold pharmacological treatment to patients with T2DM if they may need it in addition to a diet and lifestyle intervention. In addition, due to the progressive nature of T2DM, treatment intensification or switching, and non-adherence to randomised therapy would be likely to occur.

Longer follow up and large sample sizes may be observed in routinely collected primary care records, but treatment initiation will not be randomised, so initiators and non-initiators of metformin through time are likely to have differing underlying risks of cancer. By using MSMs with inverse probability of treatment weights (IPTW) and inverse probability of censoring weights (IPCW), the issues surrounding time-dependent confounding and loss to follow up can be addressed as outlined in 4.4.1. In theory, providing all assumptions are satisfied, it will be possible to compare the risk of cancer under the scenario in which everyone receives metformin monotherapy, vs the scenario where all patients are advised to follow a diet and lifestyle intervention only, with all patients adhering to their treatment and no one lost to follow up. The main objective of this chapter is the implementation of MSMs with IPTW (and IPCW) to estimate

the causal effect of metformin use and risk of cancer in patients with T2DM. This will be done using the CPRD study population identified in section 5.2. Specific further objectives include (i) the evaluation of the impact of adjusting for time-dependent confounders using MSMs vs standard methods, and (ii) careful consideration of the limitations of the methodology in the specific context of metformin use and cancer in T2DM, and due to the use of EHR data.

7.2 METHODS

A cohort of patients from the CPRD (see 5.1), with incident T2DM diagnosed between 1990 and 2014 and aged between 30 and 90 at time of diagnosis were identified, as described in section 5.2. This cohort were used as the base cohort for this study, however additional exclusions were made to enable implementation of the methods, and are detailed where relevant within this section. As explained in 4.4.2.1, study entry was defined as the beginning of the first interval after complete data on all covariates were obtained. Patients were included only if they were (i) still untreated to this point and (ii) would have still been in the risk set had they had complete data from time of diabetes diagnosis. Patients were excluded if they had a history of any cancer at study entry.

7.2.1 Exposure and comparison group definition

The exposure of interest was metformin monotherapy for first line treatment of T2DM. To avoid using a comparator that may itself have an impact on cancer risk, the comparator group were patients with a diagnosis for T2DM who were not taking any pharmacological therapy. Many of these patients had clinical codes such as “diabetic on diet only” and so it was considered reasonable to assume that they were attempting to control their disease with a lifestyle intervention, referred to going forward as “diet only”. A patient’s record in CPRD was searched for the date of first prescription for metformin, which was then taken as the date at which a patient became exposed to metformin. The BNF code used to define metformin was the code 6.1.2.2 for biguanides. Drugs that had this code in combination with another type of drug were excluded as they were indicative of exposure to dual therapy. The exposure status of an individual could update once at the date of the first metformin prescription, from diet to metformin. Further prescriptions were not needed to confirm or maintain exposure status. A patient’s follow-up was censored at subsequent use of any other medication class. This simple

definition closely mimics what would happen in a randomised controlled trial analysed by the ITT principle. Dosage and minimum adherence were not considered.

7.2.2 Outcome definition

The primary outcome was incidence of any cancer, as defined by clinical codes found in the patient's CPRD record only. The algorithm used to identify incident cancers was developed for previous work by Bhaskaran et al.[195]. A flow chart to describe an overview of this process is given in appendix 11. Briefly, in the initial stage, string searches identified codes containing key strings including (but not only) "MELANOMA" "TUMOUR" "CANCER" "MALIG" and "NEOP", as well as strings indicating treatment for cancer such as "CHEMOTHERAPY" and "RADIOTHERAPY". All codes containing the keyword strings were identified as possible cancers. A second set of key strings indicating exclusion were also used to identify which of these possible cancer codes should be excluded. Such key words here included "BENIGN" "H/O", and "SCREEN". After this stage, manual checking of the codes was used to class the codes into different groups, indicating whether the code was definite cancer, indicative, suspected, or not cancer, including whether the site was known or unknown. Further classifications grouped codes relating to treatment for cancer. All codes were then manually linked to ICD 10 to enable easy sorting into specific cancer types by a well-known and accepted coding system. Only codes that came under chapter C of the ICD coding system, classed as definite malignancies, were used to identify cases of cancer in this analysis.

For the primary analysis, the main outcome was incidence of any cancer, including non-melanoma skin cancer (NMSC). This was included to increase power, and because there is no reason to think that the mechanism by which metformin may affect cancer risk (if any) would be any different for this less serious cancer outcome. However, since it is often excluded from composite outcomes since some of the risk factors are not shared with other cancer types, a sensitivity analysis excluding NMSC was also performed. Secondary analysis also investigated the specific outcomes of lung, breast and prostate cancer as these were the three most commonly occurring cancers. Pancreatic cancer was also investigated in isolation due to its observed strong association with T2DM.

Cancer can take a long period of time to develop, and it should be considered that early stage cancer, before it has been diagnosed, may affect control of diabetes and hence result in need for treatment initiation or intensification (which results in censoring). To avoid issues of reverse

causation, and also informative censoring, the date of cancer diagnosis was brought forward in time by 6 months in the primary analysis. This was varied to 0 months and 12 months in sensitivity analyses.

7.2.3 Covariates

A list of confounders of the association between metformin use and risk of cancer, both time fixed and time- dependent were identified based on a priori knowledge and discussions with clinicians. These were included in the weighting models (and outcome models if baseline covariates) regardless of the observed associations in the data. The covariates considered in this analysis and details of how they were defined in data are outlined below. Full codelists (where applicable) and further details of how the covariates were obtained are given in appendix 12.

7.2.3.1 Time invariant confounders

Age and gender are important risk factors for cancer, and may also impact treatment decisions. Age was categorised as <45, 45-59, 60-75, > 75. Continuous non linear forms for age were also considered as detailed in 7.2.7.2.1.

Calendar time, although not causally related to the risk of cancer, is associated with the detection of cancer due to changes in screening policies etc. It could also capture some aspects of lifestyle that cannot be measured directly due to changes in knowledge through time in terms of cancer prevention. Since it is not feasible to make adjustment for the exact mediating factors, an adjustment for calendar time is needed. Calendar time is also strongly associated with treatment choice due to changes in prescribing guidelines over the study period. Calendar year of diabetes diagnosis was calculated as described in 5.2.3, and parameterised as a categorical variable of pre 1995, 1995-2000, 2000-2005 and 2005 onwards. Continuous forms of calendar year of diagnosis were not considered due to concerns over positivity violations and small sample size pre-2000.

For simplicity, smoking and alcohol consumption were considered as time invariant for the purposes of this analysis. Both variables were considered important to include due to known association with risk of cancer. These behaviours are also indicative of general unhealthy behaviour and therefore may lead a clinician to make judgements about how likely a patient is to adhere to a diet/lifestyle intervention, altering the probability of receiving pharmacological treatment.

Smoking status was categorised into non-smoker, ex-smoker, current smoker or unknown. Alcohol was defined in more detail, categorising into non-drinker, ex-drinker, current drinker unknown quantity, rare drinker <2u/d, moderate drinker 3-6u/d, excessive drinker >6u/d, and drinking status unknown. Each patient was classified into one of these categories using algorithms previously developed within the research group. Further details are given in appendix 12.

7.2.3.2 Time-dependent confounders

7.2.3.2.1 HbA1c

HbA1c is strongly associated with treatment choice, via treatment guidelines [13]. Literature also exists suggesting that HbA1c may be associated with cancer risk, though the exact mechanisms are still unknown [19, 20] . The details of how HbA1c was extracted are detailed in appendix 12. The baseline HbA1c was taken to be the closest measure prior to study entry. For patients entering the study at time of diabetes diagnosis, the HbA1c was required to have been within 6 months otherwise it was considered missing. Values below and above the 1st and 99th percentiles of the distribution were truncated by setting them to the 1st and 99th percentile values. HbA1c was categorised as <6%, 6-6.5%, 6.5-7%, 7-8%, 8-10% and >10%. As with age, non-linear continuous forms were also investigated (see 7.2.7.2.1 and7.2.7.2.2).

7.2.3.2.2 BMI

Existing literature supports the hypothesis that BMI is associated with risk of several cancers [195, 196]. Due to the potential weight reducing properties of metformin, there is also reason to believe that higher BMI will increase the probability of treatment. How BMI was extracted from CPRD is detailed in appendices 9 and 12. The cut of off 6 months prior to time of diagnosis for a valid measure was also applied for BMI. BMI changes more slowly than HbA1c, and other studies have suggested that a BMI measures as much as 3 years before would be broadly representative of the current value [195]. However, as obesity is so strongly associated with development of T2DM, it was felt that a shorter time window for valid measurement should be applied than for the general population. As with HbA1c, values were truncated at the 1st and 99th percentiles. BMI was categorised as <25, 25-29, 30-35 and > 35, and again, non-linear continuous forms were also investigated.

7.2.3.2.3 Key comorbidities

Chronic kidney disease (CKD), defined by a chronic decline in kidney function, is a common co-morbidity in patients with T2DM, and metformin is contraindicated for patients with severe chronic kidney disease [15]. The causal pathways by which CKD may be associated with cancer are unclear, however review articles summarising evidence for the association suggest one reason may be the presence of excess toxins in the body causing both CKD and cancer [197]. As such, it was considered important to adjust for CKD. In order to create a simple measure of whether a patient may have decreased kidney function the quality outcome framework (QOF) [198] preferred and accepted codes for stages of chronic kidney disease were used as a set of Read codes to identify patients with CKD and are listed in appendix 12. A patient was considered to have no CKD until the time at which the first code suggesting CKD appeared, and at this point, presence of CKD was updated to reflect this. A separate variable was also created to represent the value at baseline. The binary variable combined stages 3, 4 and 5 together. Initial investigations suggested that the number of patients initiating metformin with stage 4 or 5 CKD was too low to allow a finer categorisation than this (see appendix 13).

Presence of CVD was defined using the QOF preferred and accepted codes [198] for ischaemic heart disease, acute myocardial infarction, all types of stroke and transient ischemic attack (TIA). The full code list is given in appendix 12 .Adjustment for CVD was considered important since presence of these comorbidities will be strongly associated with diabetes severity (and as such treatment) and cancer risk due to many shared risk factors [199]. The date of first record of any CVD was recorded and a patient's record time updated from time of diagnosis onwards to reflect presence or no presence of CVD as well as a separate record to identify if it was already present at baseline.

7.2.3.2.4 Concomitant medications

To further attempt to adjust for the underlying health of the patient, which may affect risk of cancer via pathways which cannot be measured directly, information on prescriptions of other kinds of medication were obtained and adjusted for. Medications considered were:

- Statins and anti-hypertensives (Anti HTs), which will provide more detail on severity of CVD or diabetes (e.g. Presence of diabetic nephropathy). Additionally, hypertension is argued to be a risk factor for cancer [200] and ongoing use of an anti-hypertensive is an indicator for

this which is simpler and more complete than using longitudinal measures of blood pressure which may be poorly updated.

- Non-steroidal anti-inflammatory's (NSAIDs), which are indicative of chronic pain that may be caused by diabetic complications e.g. neuropathy. This may give further detail on likely severity of diabetes on top of HbA1c measures, to better control for differing cancer risks between those initiating and not initiating treatment through time.

Use of these three medications was established by searching the prescription data for BNF codes relating to those three medications (see appendix 12). For each patient, longitudinal data were collected consisting of the dates of prescriptions for each medication type for their whole history. This could then be used to calculate variables indicating use in the year prior to study entry. Equivalent time updated variables for use in the previous year were also created.

7.2.4 Interval Set up

With CPRD data, the visit schedule is defined by an individuals need to visit their GP. Different aspects of a patients care or medical history that may be relevant to a particular treatment decision may not be recorded at the same time. For example, a patient with type 2 diabetes may visit their GP and have their HbA1c measured and BMI checked. The BMI can go into the system straight away, however the HbA1c result may not be entered until the result is returned, perhaps 7 – 14 days later. Details on history of other comorbidities such as cardiovascular disease or kidney function may have been recorded at a completely different visit. Because of the unbalanced nature of the data, appropriate decisions are needed to define interval length, and to assign variable measurements into these intervals to ensure the data used to predict treatment or outcome in a given interval are appropriate.

7.2.4.1 Interval length

The results in chapter 6 showed that the key time-dependent confounders of HbA1c and BMI were updated roughly every 6 months on average. However, HbA1c is a measure of glucose control over approximately 3 months, and some patients have more frequent measures. Further to this, in order to establish temporality between predictors of treatment and treatment itself, it is conventional to use covariates measured in the previous interval to predict treatment in the current interval, as outlined in 4.4.2.2. Because of this, if we choose an interval of three months,

the predictor may have been measured up to 6 months before. Such a measure may not accurately reflect the HbA1c level influencing a patient's cancer risk when the treatment decision was made. Because of this, one month intervals were chosen for the primary analysis, with three months used in a sensitivity analysis.

7.2.4.2 Assigning variables to intervals

For intervals in which BMI or HbA1c records did not occur, last one carried forward (LOCF) was used. This assumes that this most recent value is the most representative of the true value during this time. If the reason for the missing value is that a patient has not visited the GP at all during that month, then no other time-dependent covariates will change either and nor will the patient's exposure status. Given that normal patient behaviour would be to visit the GP if any major changes occurred relating to diabetes severity or other conditions, it could be reasonably assumed that patients not visiting are stable in terms of their key covariates. If the patient has visited the GP (as indicated by changes in other covariates/exposure in this interval) then either, the measure was not updated because it has not changed, or the GP is making decisions based on the latest available data. Under either of these scenarios, LOCF seems reasonable, and is unlikely to cause serious bias in the calculation of the weights.

Due to the likely lag between measurement and recording of HbA1c, it was decided that an HbA1c entered into a patient's record would be considered to have actually occurred 7 days before, based on discussions with a GP advisor. The same was done for CKD diagnoses, as, if not already there at time of diagnoses as a "history of" code, these are likely to be in response to an eGFR or creatinine test, or a hospital diagnoses that may take time to reach the GP.

7.2.5 Including patients treated from study entry

As explained in 4.4.2.1; in order to ensure treatment has not already affected covariates, patients were required to be treatment free at the study entry. For patients that entered the study at their the time of diabetes diagnosis, the decision was made to allow patients who initiated metformin within their baseline interval to be included as long as all baseline covariates were measured strictly before the date of treatment initiation. These patients would always have a probability of treatment of 1, and so had a constant weight of 1 in the outcome model. These patients did not contribute to the models for the IPTW. Difference in baseline covariates

between these patients and those who are untreated at baseline would be adjusted for since all baseline covariates are included in the outcome model. Figure 7.1 shows inclusion/exclusion decisions for six example patients to demonstrate this, and to re-iterate how the ordering of complete data and treatment must occur for a patient to be eligible to enter the study.

Sensitivity analyses were conducted to look at the impact of including/excluding a) patients who had complete data and started medication at time of diagnosis; but whose baseline covariates occurred after the time of medication, and b) patients who started medication in the same interval in which they obtained complete data where study entry was after diabetes diagnosis.

7.2.6 Censoring

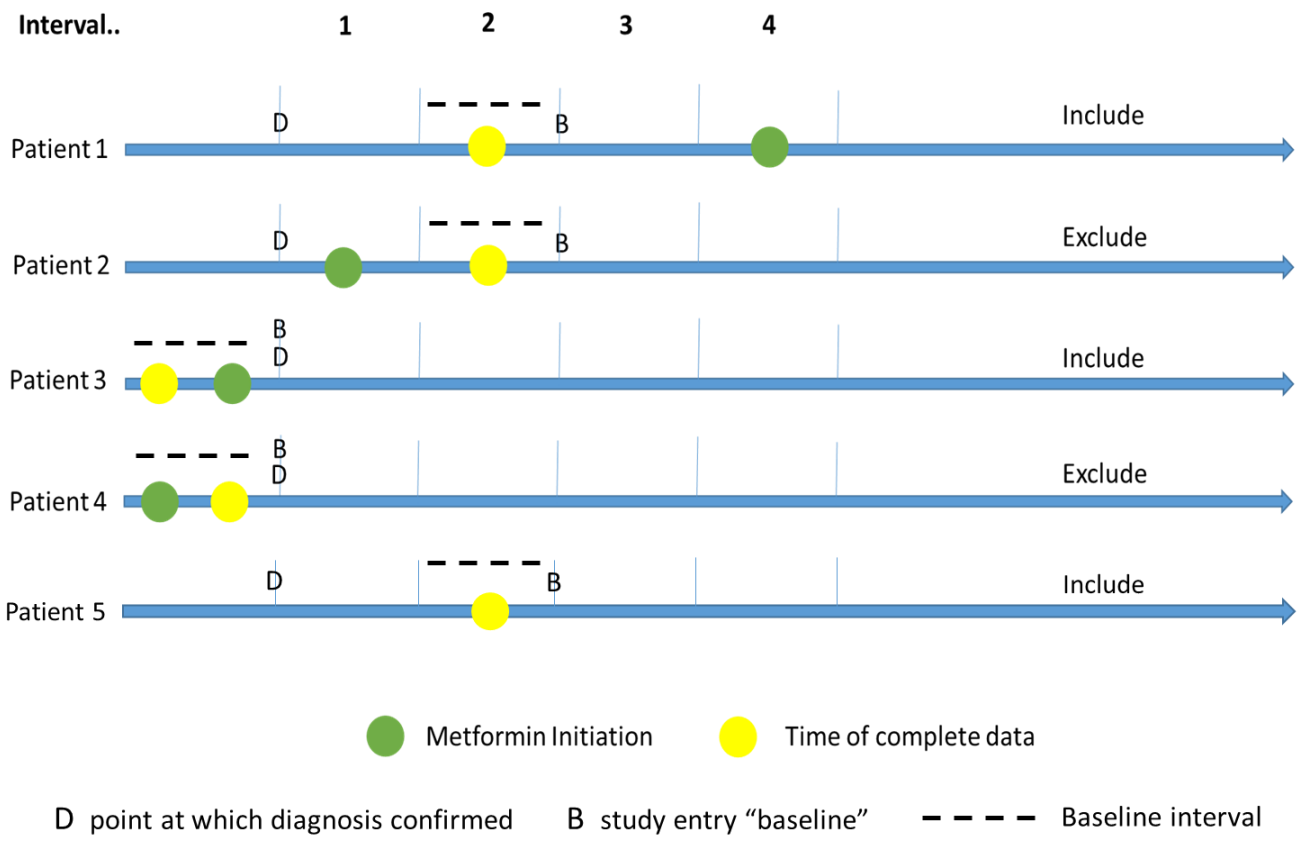
Patients were censored from follow up for the following reasons:

1. Death
2. Transfer out of practice
3. Initiation of any other diabetes medication other than metformin
4. The last data collection date.

Otherwise, follow up ended in the interval in which there was a cancer diagnosis.

The last data collection date was considered administrative and highly unlikely to be informative in terms of differing risk of outcome, therefore this was not included as a censoring event in the models for the IPCW. As explained in the methods, if a patient was censored in an interval (for reasons 1-3), they were censored from the beginning of that interval. It was felt that minimal cancer events would be missed by this, because it is unlikely that a patient would have no record of the cancer diagnoses before they died, particularly since the date of cancer diagnosis was moved forwards in time by 6 months as explained in 7.2.2.

Figure 7.1 Follow up for some example patients to show different scenarios in which a patient may or may not be included in the analysis.



Patient 1: Starts follow up in interval 3, the interval after they have complete data. Since up to this point they have not initiated treatment, then covariates have not been affected by prior medication so patient one can be included. Interval 3 counts as unexposed time, patient is exposed from interval 4.

Patient 2: Already on medication at study entry. The values contributing to the "baseline adjustment" (as defined by covariate values in interval 2) may have been influenced by treatment in interval 1 Therefore patient 2 must be excluded.

Patients 3 and 4: Initiate medication at time of diagnosis. Here we allow more leniency in that the covariates can occur in the same interval, as long as they are strictly before the time of treatment (satisfied by patient 3 but not 4). Patient 3 would be classified as being on metformin at study entry.

Patient 5: Same as patient 1 but does not then subsequently initiate treatment. This patient is included and contributes to follow up as unexposed.

7.2.7 Analysis plan

7.2.7.1 Descriptive analysis

Patient demographics (age, year of diagnosis, gender etc.) and all potential confounders were summarised overall and by treatment choice at study entry (diet only, metformin). Continuous variables were summarised as mean, SD, median and IQR. Binary/categorical variables were summarised as n (%). Crude cancer rates were summarised, including descriptive information on different cancer types. This was also useful to inform which site-specific analyses were feasible to perform.

To understand the extent to which the covariates specified in 7.2.3 were predictive of the outcome within the cohort, associations were estimated for each covariate in turn. This was performed after the data had been split into intervals, and so was done via pooled logistic regression with adjustment for time since study entry to approximate a Cox proportional hazards (PH) model. This was useful to understand in advance the extent to which the IPTW and IPCW models might be expected to produce differing results to standard statistical methods, but did not alter any decisions on covariate inclusion. To ensure that any observed associations were not solely mediated by use of metformin, or confounded by age, these analyses were adjusted for age and time updated diabetes medication (none or metformin).

7.2.7.2 Models for the IPTW and IPCW

Although the covariate selection process for this analysis was based on a priori knowledge and discussions with experts, the functional form of the covariates was not pre-specified. As outlined in 4.4.2.4, the starting point was to have two model specifications, one using natural cubic splines to model non-linear forms of association between continuous variables and log odds of cancer, and another where all variables except time were categorical.

For all model specifications, time since study entry was used as the underlying time scale, with time between diabetes diagnosis and study entry also entered as an adjustment for those patients entering later due to incomplete data. These variables were modelled as cubic splines in both model specifications.

For this analysis, three model specifications were eventually used for the models for the IPTW. Namely:

A) Time, time between diagnosis and study entry, age, HbA1c (baseline and time updated) and BMI (baseline and time updated) fitted as natural cubic splines, all other covariates categorical.

B) Simplification of the specified forms in A, to assess the sensitivity of the models to overfitting.

C) All covariates (except time and time between diagnosis and study entry) were entered as categorical variables, with the categorisations as described in 7.2.3.

7.2.7.2.1 Estimating spline forms for specification A

Natural Cubic splines were used in order to fit flexible non-linear relationships between exposure and outcome [201]. The Stata functions `UVRS` and `MVRS` [202] use an iterative based approach to choose the optimal number and location of knots, to best model the observed association between the exposure and outcome, allowing this to be done for multiple variables simultaneously. This is a data driven approach, and allowing such flexible associations could result in extreme overfitting, leading to uninterpretable "wiggles" in the curve. However, since the aim of IPTW is essentially to remove the association between treatment and time-dependent confounder within the data set, evaluating the shape of the associations as indicated solely by the data was considered a good starting point.

This function was used to calculate natural cubic splines for all continuous covariates listed for specification A above. To do this, the function was called to fit a logistic regression model to estimate probability of being treated with metformin vs diet only. This model included all continuous and categorical variables both baseline and time-dependent. Once these initial estimated spline curves had been generated, the shapes of the association between each continuous covariate and probability of treatment were plotted separately.

7.2.7.2.2 Simplification to obtain specification B

Model specification B was obtained by simplifying, where possible, the form in A, for each continuous variable. This was done with the aim of reducing issues of positivity that may occur with overfitting, and to reduce the number of parameters needed to model the association with sufficient detail. Initially, the forms from model specification A were examined for plausibility in discussion with clinicians. In general, the simplification was approached in a subjective manner, with a trial and error approach to obtaining simpler splines or polynomial forms that appeared

similar to the overall shape to those estimated by model specification A. As outlined in 4.4.2.3, decisions on whether associations could be smoothed/simplified were based on how strong the association with outcome was expected to be over the covariate range that would be affected. Decisions on whether the association with outcome over a given interval of the covariate were important were based on the descriptive univariate analyses (7.2.7.1). If the categorisations used in these analyses did not allow the relative risk of cancer over the range of interest to be estimated in enough detail, the association with cancer was re-estimated using a continuous form for the covariate, by fitting it as a natural cubic spline with four evenly spaced percentile knots (20^{th}, $40^{th,}$ 60^{th} and 80^{th}).

7.2.7.2.3 Weight calculation and truncation

The stabilised IPTW were then calculated as described in 4.4.1.2, using each of the three specifications. Due the number of potential interactions between covariates that could be tested, it was decided that interactions would not be included in the weighting models. With the exception of calendar time, there were no clear a priori covariates that were considered to be likely to modify the effect of other covariates on probability of treatment. The effect of potential interactions with calendar time were investigated in a sensitivity analysis (see 7.2.7.4).

Once the IPTW were estimated, the distribution of the weights was examined as described in 4.4.2.4, and truncation performed if necessary to obtain a mean of the IPTW close to one.

7.2.7.2.4 Models for the IPCW

IPCW were estimated as described in 4.4.1.3, using a multinomial logistic regression to model the probabilities for censoring due to death, transfer from practice, or initiation of any other medication other than metformin. The same predictors that were used for the treatment model were all included in the censoring model. For model specification A and C, the same spline parameterisations and categorisations were used for all variables. Use of the same spline forms was considered appropriate because with the exception of BMI in the previous interval, all spline parameterisations allowed flexible non-linear associations, and so would likely be sufficient to model any non linear associations with probability of censoring from medication change, death or transfer out. Any major bias resulting from incorrect specification of previous BMI in the censoring models would most likely be highlighted by comparing the results from specifications

A and C, since the categorisation of BMI would, to an extent, be expected to pick up a non linear association with censoring. Specification B was not used as it was felt that the simplified form of these splines was unlikely to be able to correctly model all associations with different reasons for censoring.

Once the IPCW were estimated, the joint weight was calculated as described in 4.4.2.2. Specifically, the untruncated IPTW was multiplied by the untruncated IPCW, and the distribution of the joint weight examined and truncated as described previously (see section 4.4.2.4).

7.2.7.3 Outcome models

MSMs also require the assumption of correct specification of the outcome model as so each model was fitted using each of the three covariate specifications described previously (A, B and C). The knot points and categorisations used were the same as in the weighting models. This was done to simplify code as far as possible during the computation process. The same spline parameterisation does not enforce the same shape of association as the treatment model, but it if the parametrisation is too simple for the association between the covariate and cancer, there is the risk of misspecifying the outcome model. To check whether this may be the case, the associations between continuous baseline covariates and outcome were plotted using covariate forms in A and B, and also using splines with 4 knots based at even percentiles in an unweighted model. No appreciable differences in shape of association were observed (see appendix 14).

For the primary analysis, exposure history was modelled using a binary variable to represent whether the patient was on/off metformin. This assumes that the effect of treatment is fully accounted for by current treatment, analogous to the single HR obtained from a Cox model. If the effect of treatment is not constant through time, then this may not be appropriate, and so cumulative use of medication was also modelled in a secondary analysis (see 7.2.7.5.1).

For each model specification, five models were fitted to evaluate the effect of metformin use vs diet on risk of cancer in patients with T2DM. For all models, the effect of metformin use on risk of cancer was estimating as described in 4.4.1.4, using a pooled logistic regression with time since study entry as the main time scale. The first 3 models were standard analyses models with varying levels of confounder adjustment. Performing standard analyses provided a comparator, so that the impact of controlling for time-dependent confounding using MSMs in the later models could be investigated. Although it would be possible to fit Cox models for the first three

models, rather than the approximation by a pooled logistic regression model, it was decided for simplicity and ease of comparability to use a pooled logistic regression for all models. The 4th and 5th models were marginal structural models with IPTW and then joint IPTW and IPCW. The exact specifications of each model are outlined below:

Model 1 – Minimal adjustment for confounding: adjustment for age, gender, smoking status and alcohol status and year of onset of diabetes.

Model 2 – Full adjustment for baseline covariates: Model 1 + baseline adjustment for: HbA1c, BMI, use of other medications in previous year (NSAIDS, statins, antihypertensive drugs), history of chronic kidney disease (CKD) and cardiovascular disease (CVD).

Model 3 – Full adjustment for baseline covariates with time-dependent covariates added: Models 2 + adjustment for time updated HbA1c, BMI, and history of CVD, CKD and use of other medications in the past 12 months.

Model 4 – As model 2, weighted using IPTW. (MSM with IPTW)

Model 5 – As Model 2, weighted using joint IPTW and IPCW (MSM with IPTW and IPCW)

It was not planned to explore interactions between treatment and baseline covariates in the main analysis, as there was no clear reason to think they would exist. Additionally, the overall aim was to estimate the effect of metformin in the general diabetic population and therefore specific interactions were not of interest.

7.2.7.4 Sensitivity analyses

To examine how various modelling decisions may have affected the results of the MSMs, the treatment and censoring weights were re-calculated (if required) and the outcome models re-run for nine scenarios (listed below) as sensitivity analyses. All sensitivity analyses were conducted for each of the three covariate specifications A, B and C.

7.2.7.4.1 Dividing time into 3-month instead of 1-month intervals

The descriptive analysis presented in chapter 6 suggested that a patient's HbA1c and BMI were updated on average just over every 6 months. It also suggested that measures were commonly updated within 1 month of treatment. This sensitivity analysis was conducted to investigate whether the results could be altered if larger intervals were used.

7.2.7.4.2 Different lags on the cancer diagnosis

In the primary analysis, the date of cancer diagnosis was moved forward in time by 6 months to allow for un-diagnosed cancer. This was because developing cancer may affect diabetes symptoms and therefore affect treatment decisions. To test the sensitivity of the models to this assumption (i.e. to assess the potential impact of reverse causation on results), two sensitivity analyses were performed, one where the cancer diagnosis date was shifted forward by 12 months, and one where it was not changed at all.

7.2.7.4.3 Excluding non-melanoma skin cancer from outcome

As explained in 7.2.2, non melanoma skin cancer was included in the definition of "any cancer" since it improved power and there was no reason to think the mechanism by which metformin may affect risk was any different to other cancers. However, since it is often excluded from definitions of all cancer, the analysis was re-run excluding this cancer type. This was done by excluding it entirely from the definition of cancer. Specifically, any previous history of NMSC was ignored, and if a patient developed, NMSC during follow up, this was ignored and other cancers occurring after this were included as an incident primary cancer.

7.2.7.4.4 Assigning a patient as exposed from 1 year after first prescription

Assuming continuous exposure to metformin after the first prescription, this sensitivity analysis aimed to investigate whether the relative risk of cancer with metformin use was any different if outcomes observed within 1 year of starting metformin were not attributed to the "exposed" group. This, to an extent, tested the sensitivity of the results to making differing assumptions about how quickly any effect of metformin may become apparent. This was investigated further in a secondary analysis looking at cumulative medication use.

7.2.7.4.5 Relaxing assumptions of temporality between covariate measurement and treatment

In the primary analysis, patients who initiated treatment in the same interval in which they obtained complete data were excluded. This group of patients could be divided into two distinct groups: firstly, patients initiating at time of diagnosis whose baseline covariates were measured

after treatment initiation; and secondly, patients who initiated treatment post baseline but in the same interval in which they obtained complete data. These two sensitivity analyses looked at the impact of including these two groups of patients.

7.2.7.4.6 Using covariates in current interval to predict treatment in weight calculation

By using covariates from the previous interval to predict treatment initiation and censoring (as is standard practice for IPW), we may not have correctly controlled for confounding at the time the treatment decision was actually made. This is because for many covariates, the true value may be measured at a GP consultation and treatment initiated (or not) immediately in response to this. By using covariates from the same interval to predict treatment initiation, it was possible to investigate whether the potential to be missing the true value that represents outcome risk and drives the treatment decisions may have influenced the estimated relative risks.

7.2.7.4.7 Fitting different treatment models for each calendar period

In the overall model fitting process, potential interactions between covariates were not examined. Due to changing guidelines, there was some concern that calendar time may have been an important effect modifier for the effect of covariates on probability of treatment. For example, in the 1990s, those patients who were not overweight were more likely to be prescribed a sulfonylurea than metformin, however more recent guidelines indicate that metformin should be used in preference. Therefore, the effect of BMI on probability of treatment may be modified by calendar time. To control for this, IPTW and IPCW were estimated separately for patients diagnosed in different calendar periods. Due to small numbers for some time intervals, this analysis was restricted to a smaller range of calendar periods, including only patients diagnosed from the year 2000 onwards.

7.2.7.5 Secondary analyses

7.2.7.5.1 Modelling cumulative use of metformin

One of the biggest anticipated limitations of this analysis is the probable lack of follow up time in which to observe an effect of metformin use on cancer risk if it exists. In line with the majority

of studies that have previously examined this association, the primary analysis assumed a constant effect of treatment through time. However, in reality, it is likely that an effect of metformin may take years to become apparent. By instead defining treatment as a categorical variable representing time since first metformin prescription, the effect of longer term metformin use was estimated. Time since start of metformin was categorised as <6 months, 6-12 months, 1-2 years, 2-5 years and 5-7 years and > 7 years. As with the primary analysis, this was estimated for all three covariate specifications.

7.2.7.5.2 Site specific analysis

The three most common site specific cancers (namely breast, lung, prostate) were analysed as separate, secondary outcomes, to investigate whether the risk of cancer with metformin use may differ between cancer types. In addition, pancreatic cancer was also investigated. There is a strong association between risk of pancreatic cancer and type 2 diabetes, so it was hypothesised that the issues of time-dependent confounding affected by prior treatment may be stronger for this cancer type.

Since having any cancer can modify the risk of subsequent cancers, patients who were diagnosed with a cancer other than the type of interest were censored. Therefore, another category level for censoring was added in this analysis, and new IPCW estimated for each cancer type of interest. IPTW were also re-estimated for the relevant populations for each cancer type. E.g. breast and prostate cancer analyses were restricted to women and men respectively. As with the primary analysis, the site specific analyses were run with the 3 different covariate specifications, and all cancer diagnoses dates brought forward in time by 6 months.

7.2.7.5.3 Sulfonylurea vs diet

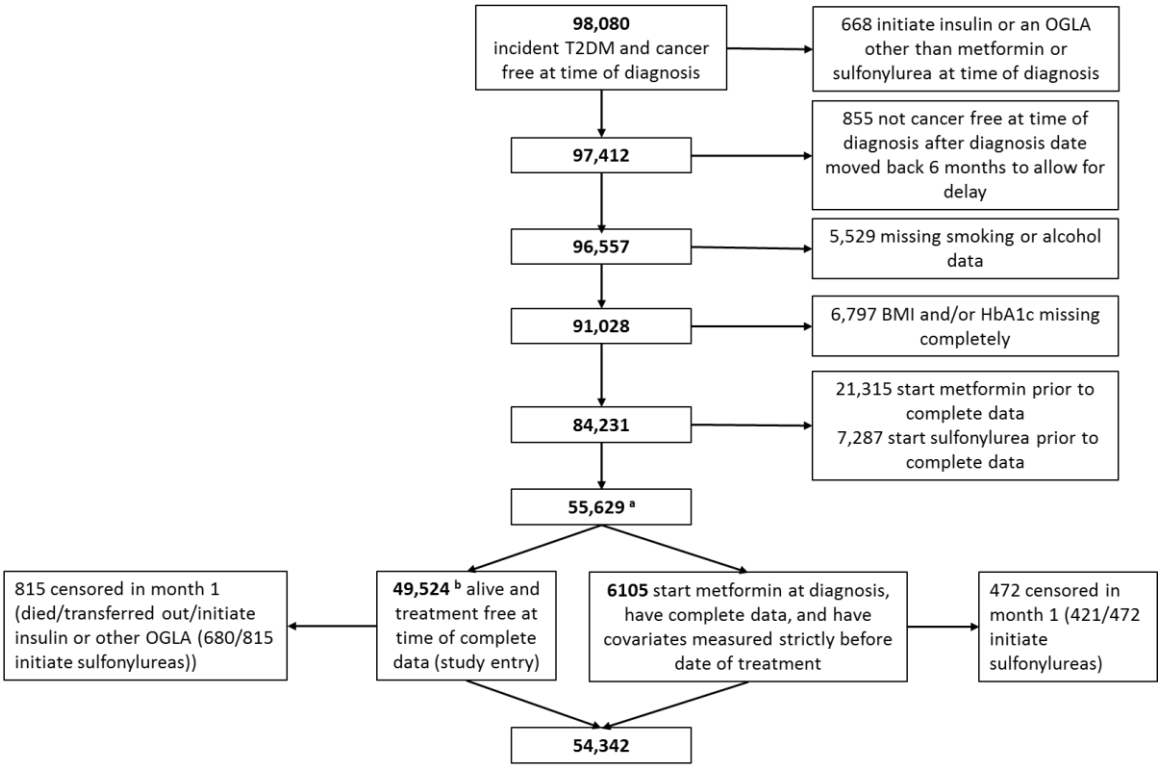
Risk of cancer was also compared for use of sulfonylureas vs diet only. This comparison allowed informal inferences on how the risk of cancer may differ between metformin and sulfonylurea users, which enabled comparison of these results to many of the existing observational studies that used active comparator designs. Sulfonylurea use was defined and recorded in the same way as metformin use in the primary analysis, but using the BNF code 6.1.2.1. Patients were censored if they initiated any other drug than a sulfonylurea. IPTW and IPCW were estimated using the same approach as for the primary analysis.

7.3 RESULTS

7.3.1 Basic cohort description

From the 98,080 patients identified as having incident type 2 diabetes and who were free of any cancer at time of diabetes diagnosis in section 5.2, a total of 55,629 were alive, cancer free and eligible to enter the study. Of these, 54,342 were not censored in month one due to death, transfer out of practice, or initiation of another therapy, and contributed to the outcome model. Figure 7.2 shows how this cohort was obtained from the original population.

Figure 7.2 Flowchart to show how initial population of 98,080 were reduced into population of 54,342contributing to outcome model.



[a] 55,629 contribute to model for censoring weights.
[b] the 49,524 untreated at study entry, less the 815 censored and 48 with a cancer diagnosis in month 1 result in 48,661 patients contributing to the model for IPTW.

During follow up, just under 50% of the cohort were censored due to death, transfer out, or initiation of other therapy. The most common reason for censoring was initiation of a sulfonylurea, which could occur both before or after a patient had started metformin. Table 7.1 presents the reasons for study exit for all patients eligible for study entry (n=55,629).

Reason for end of follow up	N	%
End of study (last collected date for practice)	28,979	52.1%
Cancer event	2,530	4.5%
Initiates Insulin	1,898	3.4%
Initiates OAD other than metformin or sulfonylureas	3,964	7.1%
Initiates sulfonylureas	8,642	15.5%
Transfer out of practice	5,554	10.0%
Died	4,062	7.3%
Total	**55,629**	**100%**

Table 7.1 Proportion of patients and reasons for exiting the study.

The characteristics of all 55,629 patients alive and cancer free at study entry are displayed in Table 7.2, shown overall, and by medication at study entry. Mean follow up time overall was 3.8 years. Total time contributing to diet only was 119,325 person years, and 79,267 person years for metformin. The average age at diabetes diagnosis was 62 years (SD 12 years). Patients initiating metformin at study entry were slightly younger (mean age 58 years, SD 12 years). Mean HbA1c and BMI were higher in those initiating metformin at study entry than not, and patients with a history of chronic kidney disease were less likely to initiate metformin. Numbers initiating metformin at study entry also increased with increasing calendar time as expected due to changes in treatment guidelines.

Within the 54,342 patients who contribute to the outcome model, 2,530 had a cancer event during follow up. The crude rates of events for no medication and metformin were 13.7 per 1000 person years and 11.2 per 1000 person years respectively. Table 7.3 presents a breakdown of cancer type as defined by ICD code. Table 7.4 presents the estimated associations between all covariates and incidence of any cancer. As expected, the strongest evidence of an association was observed for age, where increasing age was associated with increasing risk of cancer. Additionally, current and ex-smokers were also estimated to have an increased risk of cancer compared to non-smokers. There was also some suggestion that compared to a reference level of <6%, all categories of HbA1c in the previous interval were associated with a higher risk of cancer, with the risk being higher for categories spanning an HbA1c of 6% -8%. The majority of covariates, both baseline and time-dependent, had estimated HRs with confidence intervals that contained one.

	No Therapy N = 49,524	Metformin N=6,105	Total N=55,629
Mean (SD) median, 25th %ile – 75th %ile			
Age at diagnosis	62.2 (12) 63 ,54 - 71	57.6 (11.8) 57 ,49 - 66	61.7 (12) 62 ,53 - 71
Time (months) to complete data	3.8 (3) 3.1 ,1.3 - 5.6	0 (0) 0 ,0 - 0	3.4 (9.7) 0 ,0 - 3
HbA1c	7.2 (1.6) 6.8 ,6.2 - 7.7	9.4 (2.3) 9 ,7.4 - 11	7.5 (1.8) 6.9 ,6.3 - 8
BMI	31.6 (6.3) 30.7 ,27.3 - 34.9	33.4 (6.9) 32.3 ,28.6 - 37.1	31.8 (6.3) 30.9 ,27.5 - 35.2
N (%)			
Gender			
Male	27763 (56.1)	3594 (58.9)	31357 (56.4)
Female	21761 (43.9)	2511 (41.1)	24272 (43.6)
History of Chronic Kidney Disease			
No	46463 (93.8)	5866 (96.1)	52329 (94.1)
Yes	3061 (6.2)	239 (3.9)	3300 (5.9)
History of Cardiovascular Disease			
No	41868 (84.5)	5479 (89.7)	47347 (85.1)
Yes	7656 (15.5)	626 (10.3)	8282 (14.9)
Use of statins in previous year			
No	25035 (50.6)	2739 (44.9)	27774 (49.9)
Yes	24489 (49.4)	3366 (55.1)	27855 (50.1)
Use of NSAID in previous year			
No	39575 (79.9)	4999 (81.9)	44574 (80.1)
Yes	9949 (20.1)	1106 (18.1)	11055 (19.9)
Use of Anti HT in previous year			
No	18048 (36.4)	2767 (45.3)	20815 (37.4)
Yes	31476 (63.6)	3338 (54.7)	34814 (62.6)
Smoking Status			
Never	20132 (40.7)	2449 (40.1)	22581 (40.6)
Current	8746 (17.7)	1287 (21.1)	10033 (18)
Ex	20646 (41.7)	2369 (38.8)	23015 (41.4)
Alcohol consumption			
non-drinker	5770 (11.7)	884 (1.8)	6654 (13.4)
ex-drinker	3474 (7)	529 (1.1)	4003 (8.1)
current drinker quantity unknown	979 (2)	121 (0.2)	1100 (2.2)
rare drinker <2u/d	11543 (23.3)	1484 (3)	13027 (26.3)
moderate drinker 3-6u	22934 (46.3)	2570 (5.2)	25504 (51.5)
excessive drinker >6u	4824 (9.7)	517 (1)	5341 (10.8)
Calendar Year of onset			
1990 - 1995	134 (0.3)	0 (0)	134 (0.2)
1995 - 2000	1708 (3.5)	20 (0.3)	1728 (3.1)
2000-2005	12764 (25.8)	595 (9.8)	13359 (24)
post 2005	34918 (70.5)	5490 (89.9)	40408 (72.6)

Table 7.2 Baseline demographics of patients eligible for study entry , by treatment at study entry (treatment at study entry defined as in 7.2.5)

ICD code/Site	Number of events	%
C43 & 44 /Malignant Skin Cancer	738	29
C61/Prostate Cancer	266	11
C50/Breast Cancer	241	10
C34/Lung Cancer	185	7
C18/Colon Cancer	158	6
C67/Bladder Cancer	69	3
C25/Pancreatic Cancer	50	2
C15/Oesophageal Cancer	65	3
C20/Rectal Cancer	68	3
C54/Endometrial Cancer	47	2
C85/ Lymphoma	40	2
Other*	603	24
Total	**2530**	**100**

Table 7.3 Frequency table of cancer types occurring in study cohort

*all other cancers with less than 40 events each, plus ICD C80 for "cancer of unspecified site"

Risk Factor	Hazard ratio	95% CI
Age		
<45	1(ref)	
45-59	3.48	(2.38 , 5.10)
60-74	8.12	(5.58 , 11.81)
75+	12.07	(8.26 , 17.64)
Gender		
Male	1 (ref)	
Female	0.75	(0.70 , 0.82)
Smoking status		
Never	1(ref)	
Current	1.25	(1.10 , 1.41)
Ex	1.27	(1.17 , 1.39)
Drinking Status		
Non drinker	1(ref)	
ex-drinker	1.08	(0.89 , 1.32)
current drinker unknown	1.26	(0.95 , 1.67)
rare drinker <2u/d	0.97	(0.84 , 1.12)
moderate drinker 3-6u/d	1.09	(0.96 , 1.24)
excessive drinker >6u/d	1.15	(0.97 , 1.37)
Year of diabetes onset		
pre 1995	1 (ref)	
1995 - 2000	0.67	(0.32 , 1.40)
2000-2005	0.71	(0.34 , 1.48)
post 2005	0.68	(0.32 , 1.42)
Any history at baseline of …		
CVD	1.12	(1.02 , 1.24)
CKD	1.08	(0.92 , 1.27)
Time updated history of…		
CVD	1.12	(1.02 , 1.22)
CKD	1.06	(0.95 , 1.18)
Use in year before baseline of…		
Anti-hypertensive medications	1.05	(0.96 , 1.15)
Statins	0.98	(0.91 , 1.06)
NSAIDS	0.99	(0.90 , 1.09)
Time updated use in previous year of…		
Anti-hypertensive medication	1.09	(0.99 , 1.21)
Statins	1.06	(0.97 , 1.16)
NSAIDS	0.93	(0.84 , 1.03)
Baseline BMI [a]		
<25	1(ref)	
25-29	0.92	(0.78 , 1.09)
30-34	1.08	(0.88 , 1.32)
35+	0.92	(0.71 , 1.20)
Baseline HbA1c [a]		
<6%	1(ref)	
6-6.5%	0.89	(0.78 , 1.01)
6.5 -7%	0.99	(0.87 , 1.14)
7-8%	0.97	(0.83 , 1.12)
8-10%	0.91	(0.76 , 1.09)
10% +	0.82	(0.65 , 1.03)
Previous BMI (interval -1) [b]		
<25	1(ref)	
25-29	0.95	(0.82 , 1.11)
30-34	0.85	(0.71 , 1.03)
35+	0.93	(0.71 , 1.20)
Previous HbA1c (interval -1) [b]		
<6%	1(ref)	
6-6.5%	1.15	(1.01 , 1.31)
6.5 -7%	1.15	(1.00 , 1.32)
7-8%	1.12	(0.96 , 1.30)
8-10%	1.07	(0.86 , 1.33)
10% +	1.06	(0.72 , 1.57)

Table 7.4 HR and 95% CI for associations between covariates and cancer. Each covariate considered in turn, Adjusted for age, and time updated diabetes medication (none/metformin).

[a] Additionally adjusted for value of covariate in previous interval [b] Additionally adjusted for baseline value of covariate

7.3.2 Primary analysis

7.3.2.1 Spline fitting

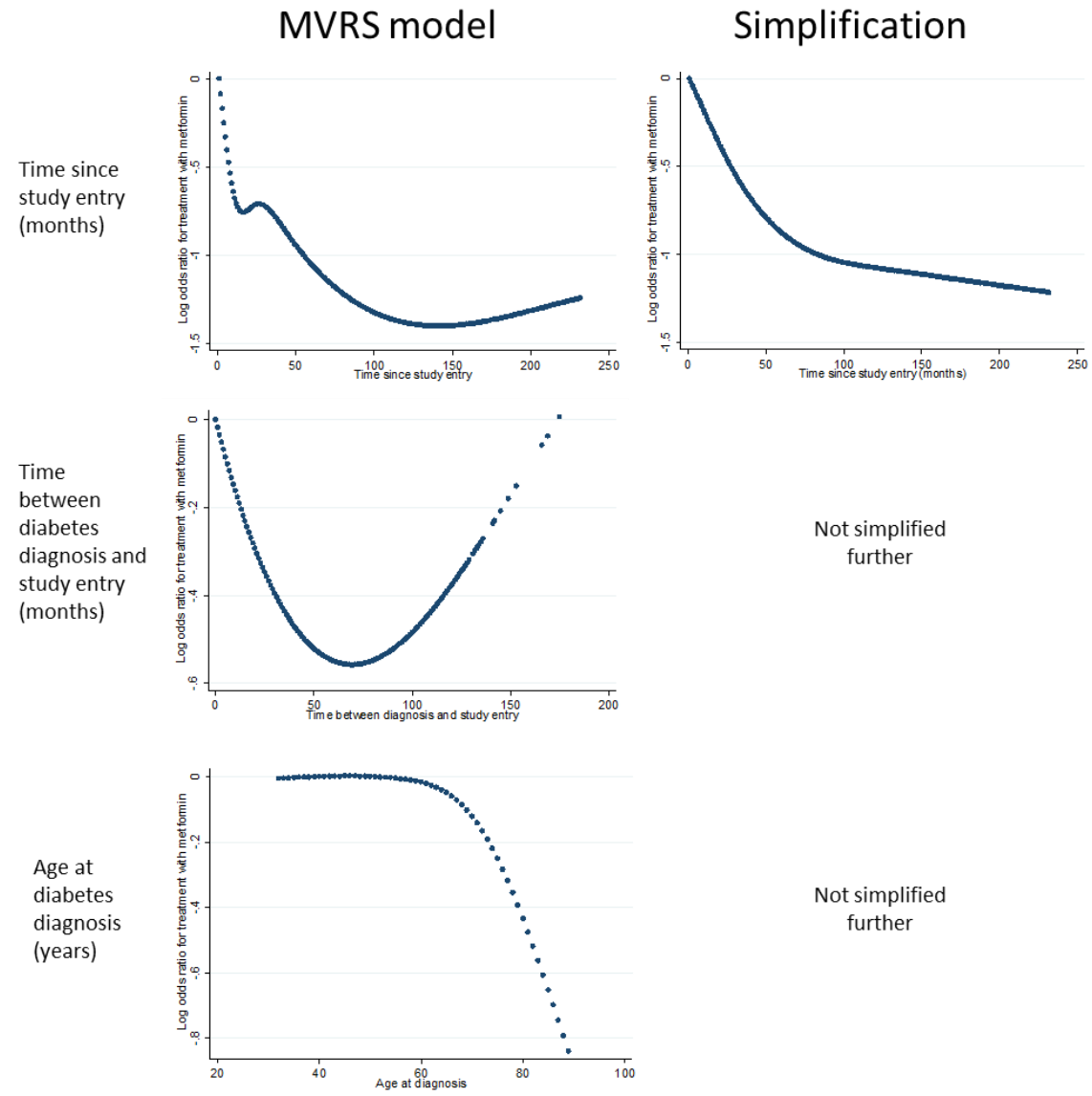
48,661 subjects not treated at study entry contributed to the model for the IPTW (see Figure 7.2 footnote b). Using the Stata function `MVRS`, natural cubic splines were fitted to estimate the probability of metformin initiation in a full multivariable model. The log odds ratios for the associations with treatment as estimated by the cubic splines are displayed in Figure 7.3 (for time since study entry, time between diagnosis and study entry, and age) and Figure 7.4 (for baseline and time updated BMI and HbA1c), alongside the simplified spline forms (if applicable) that were created as outlined in 7.2.7.2.2.

Time since study entry was simplified to remove an uninterpretable "wiggle" in the curve between 0 and 5 years from study entry. Both previous and baseline HbA1c were mostly quadratic in shape, and therefore these two curves were simplified by using a linear and quadratic term rather than a cubic spline. BMI at study entry was mostly linear, except for a slightly shallower gradient for those with BMI below about 28. This was simplified to a linear term. Spline terms for age at onset, previous BMI (which was already linear) and time between diagnosis and study entry were not simplified further from their initial spline form as chosen by MVRS function. Table 7.5 outlines the final forms for continuous covariates for each of the three model specifications A, B and C.

	A (MVRS output)	**B Simplified**	**C Categorised**
	natural cubic spline with knots at minimum and maximum plus…		
Time since study entry (months)	10 , 25 , 44	natural cubic spline with knots at knots at 10,25, 120	As B
Time between diagnosis and study entry (months)	4	as A	as A
HbA1c (%) in previous interval	6 , 6.8	linear and quadratic term	<6, [6-6.5), [6.5-7) , [7-8) , [8-10) , 10 +
BMI in previous interval	Linear	as A	<25 , [25-30), [30-35), 35+
Age at diagnosis	56 , 72	as A	32-44 , 45--59 , 60-74 , 75+
HbA1c (%) at study entry	5.9 , 6.4 , 6.9	linear and quadratic term	<6, [6-6.5), [6.5-7) , [7-8) , [8-10) , 10 +
BMI at study entry	26.8 , 30.1	linear	<25 , [25-30), [30-35), 35+

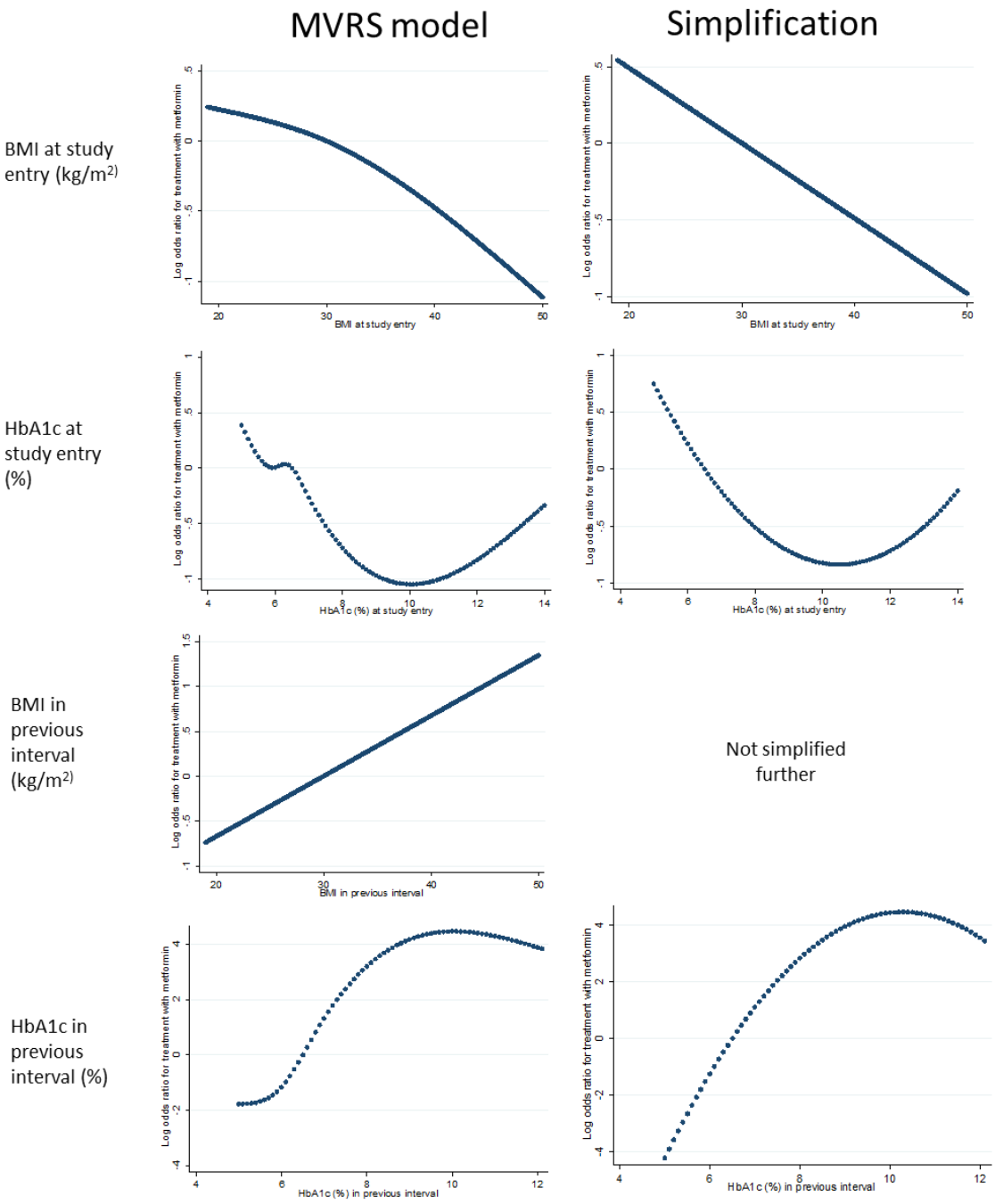
Table 7.5 Forms for continuous covariates in different model specifications.

Figure 7.3 Associations between continuous variables (time and age) and treatment with metformin from multivariable model.



Associations where spline knots and knot points decided by iterative function MVRS are shown on the left. If simplification deemed possible, this is shown on the right.
Reference time = 1, Reference age = 50.

Figure 7.4 Associations between continuous variables (BMI and HbA1c) and treatment with metformin from multivariable model.



Associations where spline knots and knot points decided by iterative function MVRS are shown on the left. If simplification deemed possible, this is shown on the right.
Reference BMI = 30 kg/m^2, Reference HbA1c = 6.5%

7.3.2.2 Models for the IPTW

Table 7.6 displays the parameter estimates for models predicting probability of treatment with metformin through time, for both the denominator and numerator models for the IPTW calculations. Results are shown for covariate specification C only. Full model outputs for A and B are displayed in appendix 15.

HbA1c was a strong predictor of treatment, with higher HbA1c in the previous interval resulting in higher odds of being treated with metformin. Very large ORs were observed for the highest values of HbA1c. This indicated that there may be some near violations of the positivity assumption for these models. In contrast, baseline HbA1c had the opposite direction of effect, suggesting that change in HbA1c from baseline was important. The same pattern was seen for BMI, though the estimated ORs were smaller in magnitude than for HbA1c. Other covariates such as age, gender, smoking, history of CKD and use of statins, anti HTs and NSAIDs were also associated with treatment, although the associations were not as large in magnitude as for HBA1c. Increasing age at diabetes diagnosis and history of CKD were associated with a decreased probability of treatment, whereas being an ex or current smoker was estimated to increase the probability of treatment compared to a non-smoker. Females were estimated to be more likely to be treated with metformin than males. All concomitant medications were estimated to increase the probability of being treated, with the estimates for use at study entry suggesting that the increased probability of treatment was related to initiation of anti-hypertensives or statins since baseline.

Table 7.7 shows the distribution of the IPTWs calculated from the models described above, including patients who were treated from baseline and therefore had a weight of one throughout follow up. Both the un-stabilised and stabilised weights are shown for comparison purposes. As expected, the un-stabilised weights had very large means and standard deviations. After stabilisation, the mean weights under the two continuous models were still extremely high.

A brief investigation was undertaken to look at the characteristics of the subjects contributing to the top 1% of the IPTWs after stabilisation (see Table 7.8). 958 unique individuals made up the top 1% of the IPTW. Of these 958, 29% were patients who were treated with an HbA1c < 6.5%. 60% were patients not treated with a previous HbA1c >=8%. Examining the distribution of covariates more generally, the mean previous HbA1c in those with the top 1% of weights who were untreated was 8.87, versus 6.45 in those untreated but in the bottom 99% of IPTW. In

those treated with metformin, the corresponding means were 7.05 and 8.15 respectively. Additionally, those treated with metformin who had a large weight, tended to be treated later in time, and their baseline HbA1c levels were relatively high (mean 8.08), suggesting that these are subjects whose IPTW increases through time due to not being treated when perhaps they should have been, and so stays high once treated. There were no clear differences in BMI, time between diagnosis and study entry, or any large differences in age at diagnosis between patients that had weights in the top 1% or bottom 99% of the distribution.

Weights were initially truncated at the 99th and 1st percentiles, which resulted in a mean weight just below 1. A slightly more lenient truncation at absolute weight values of 10 and 0.1 gave a mean weight of 1.01 for all three models (Table 7.7). The truncation at 10 and 0.1 was the one carried forward for use in the MSM with IPTW only.

	DENOMINATOR MODEL			NUMERATOR MODEL		
	OR	**SE**	**95% CI (OR)**	**OR**	**SE**	**95% CI (OR)**
BASELINE FIXED						
*Time since study entry (months) spl 1	0.98	0.001	0.98 , 0.98	0.98	0.001	0.98 , 0.98
*Time since study entry (months) spl 2	1.05	0.004	1.04 , 1.06	1.08	0.004	1.07 , 1.08
*Time between diagnosis and study entry spl 1	0.87	0.022	0.82 , 0.91	0.85	0.027	0.80 , 0.91
*Time between diagnosis and study entry spl 2	1.06	0.020	1.02 , 1.10	0.86	0.020	0.82 , 0.90
Age at diagnosis (years)						
32-44	1 (ref)			1 (ref)		
45-59	0.97	0.032	0.91 , 1.04	0.96	0.330	0.90 , 1.02
60-74	0.88	0.029	0.83 , 0.94	0.76	0.024	0.71 , 0.81
75-89	0.63	0.025	0.58 , 0.68	0.51	0.019	0.47 , 0.55
Gender (F v M)	1.13	0.019	1.09 , 1.17	1.12	0.019	1.08 , 1.15
Smoking Status						
Never	1 (ref)					
Current	1.03	0.024	0.99 , 1.08	1.10	0.025	1.06 , 1.15
Ex	1.07	0.019	1.03 , 1.10	1.07	0.019	1.04 , 1.11
Alcohol consumption						
non_drinker	1 (ref)					
ex-drinker	0.97	0.040	0.90 , 1.06	0.96	0.037	0.89 , 1.04
current drinker unknown	0.82	0.057	0.71 , 0.94	0.87	0.053	0.77 , 0.98
rare drinker <2u/d	0.98	0.028	0.93 , 1.04	0.97	0.027	0.92 , 1.03
moderate drinker 3-6u/d	0.98	0.026	0.93 , 1.03	0.94	0.024	0.89 , 0.99
excessive drinker >6u/d	0.93	0.033	0.86 , 0.99	0.87	0.030	0.81 , 0.93
Year of diabetes onset						
1990-1994	1 (ref)					
1995-2000	1.30	0.268	0.87 , 1.95	1.19	0.243	0.79 , 1.77
2001-2005	1.33	0.273	0.89 , 1.99	1.14	0.229	0.77 , 1.69
2005 onwards	1.39	0.286	0.93 , 2.08	1.11	0.224	0.75 , 1.65
Use of anti HT in year prior to study entry	0.87	0.027	0.81 , 0.92	0.97	0.017	0.93 , 1.00
Use of statin in year prior to study entry	0.89	0.021	0.85 , 0.93	1.21	0.021	1.17 , 1.25
Use of NSAID in year prior to study entry	1.09	0.025	1.04 , 1.14	1.16	0.021	1.12 , 1.20
HbA1c at study entry						
<6%	1 (ref)					
6% - 6.5%	0.86	0.031	0.80 , 0.92	1.61	0.045	1.52 , 1.70
6.5%-7%	0.76	0.028	0.71 , 0.82	2.52	0.068	2.39 , 2.66
7% - 8%	0.62	0.024	0.57 , 0.67	5.21	0.139	4.95 , 5.49
8%-10%	0.44	0.020	0.40 , 0.48	10.26	0.329	9.64 , 10.93
>10%	0.42	0.025	0.38 , 0.48	15.9	0.670	14.64 , 17.27
BMI at study entry						
<25	1 (ref)					
25-29	0.96	0.046	0.88 , 1.06	1.41	0.043	1.33 , 1.49
30-34	0.89	0.051	0.79 , 0.99	1.55	0.048	1.46 , 1.65
35+	0.76	0.052	0.66 , 0.87	1.76	0.057	1.65 , 1.87
History of CVD at study entry	0.92	0.053	0.82 , 1.03	1.03	0.023	0.99 , 1.08
History of CKD at study entry	0.82	0.047	0.74 , 0.92	0.69	0.028	0.64 , 0.75
TIME UPDATED						
Use of anti HT in previous year	1.07	0.024	1.03 , 1.12			
Use of statin in previous year	1.55	0.037	1.48 , 1.63			
Use of NSAID in previous year	1.21	0.039	1.14 , 1.29			
History of CVD	1.03	0.057	0.93 , 1.15			
History of CKD	0.87	0.036	0.80 , 0.94			
HbA1c in previous interval						
<6%	1 (ref)					
6% - 6.5%	2.35	0.141	2.09 , 2.64			
6.5%-7%	7.13	0.409	6.37 , 7.98			
7% - 8%	41.58	2.357	37.21 , 46.47			
8%-10%	186.22	11.370	165.22 , 209.89			
>10%	311.57	22.934	269.71 , 359.92			
Bmi in previous interval						
<25	1 (ref)					
25-29	1.38	0.064	1.26 , 1.52			
30-34	1.67	0.093	1.50 , 1.86			
35+	2.21	0.147	1.94 , 2.52			

Table 7.6 Estimated OR, standard error and 95% CI for probability of treatment with metformin for denominator and numerator models for the IPTW, covariate specification C.

*estimates for time since study entry and time between diagnosis and study entry equivalent to the simplified spline in Figure 7.3.

		Mean	Standard Deviation	1st %ile	5th %ile	10th %ile	25th %ile	50th %ile	75th %ile	90th %ile	95th %ile	99th %ile	Minimum	Maximum
MVRS (A)	Un -stabilised	389944	281000000	1.01	1.02	1.03	1.09	1.47	15.42	72.55	170.06	743.57	1	$2.88x10^{11}$
	Stabilised	11871	5514653	0.04	0.14	0.29	0.68	0.89	0.98	1.38	2.21	6.41	$1.23x10^{-6}$	$4.83x10^{9}$
	Stabilised Truncated at 1st and 99th %ile	0.98	0.86	0.04	0.14	0.29	0.68	0.89	0.98	1.38	2.21	6.41	0.04	6.41
	Stabilised Truncated at 0.1 and 10	1.01	1.07	0.10	0.14	0.29	0.68	0.89	0.98	1.38	2.21	6.41	0.1	10
Simplified (B)	Un -stabilised	85990	60900000	1	1.01	1.03	1.1	1.48	16.77	73.32	156.76	787.67	1	$6.23x10^{10}$
	Stabilised	3098	1364186	0.04	0.15	0.31	0.64	0.89	0.99	1.36	2.12	7.78	$3.99x10^{-8}$	$1.17x10^{9}$
	Stabilised Truncated at 1st and 99th %ile	0.99	0.98	0.04	0.15	0.31	0.64	0.89	0.99	1.36	2.12	7.78	0.04	7.78
	Stabilised Truncated at 0.1 and 10	1.01	1.13	0.10	0.15	0.31	0.64	0.89	0.99	1.36	2.12	7.78	0.10	10
Categorical (C)	Un -stabilised	104.5	14181	1	1.02	1.03	1.09	1.49	16.91	78.81	188.98	673.11	1	$1.18x10^{8}$
	Stabilised	2.34	281	0.04	0.15	0.29	0.65	0.88	0.99	1.46	2.39	6.13	$5.32x10^{-6}$	187863
	Stabilised Truncated at 1st and 99th %ile	0.98	0.85	0.04	0.15	0.29	0.65	0.88	0.99	1.46	2.39	6.13	0.04	6.13
	Stabilised Truncated at 0.1 and 10	1.01	1.06	0.1	0.15	0.29	0.65	0.88	0.99	1.46	2.39	6.13	0.1	10

Table 7.7 Distribution of inverse probability of treatment weights (unstabilised, stabilised and two different truncations) from treatment models with differing covariate specifications.

	MVRS (A)				Simplified (B)				Categorical (C)			
	No Medication		Metformin		No Medication		Metformin		No Medication		Metformin	
Top 1% of Weights												
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Stabilised Weight value	3463088	9.54E+07	1034535	2.06E+07	996745.2	2.48E+07	317304.2	6.31E+06	375.0967	4.87E+03	110.1523	1.41E+03
Previous HbA1c	8.87	1.47	7.05	1.80	9.07	1.49	7.03	2.07	8.79	1.56	6.82	1.55
Previous BMI	33.17	5.77	32.74	6.55	33.40	6.09	32.69	6.37	33.48	5.97	32.53	6.22
Age at onset	55.17	10.63	55.81	11.75	54.42	10.64	55.75	11.80	55.15	10.82	57.03	11.45
Time since study entry	51.74	26.00	28.38	23.59	49.01	27.04	26.79	24.42	51.43	26.36	27.06	24.53
Calendar Year of onset	2005	3.50	2005	4.19	2005	3.51	2005	4.14	2005	3.51	2005	4.22
Baseline BMI	33.45	5.88	33.74	6.56	33.72	6.19	33.74	6.65	33.93	6.02	33.61	6.29
Baseline HbA1c	7.34	1.72	7.98	1.66	7.60	1.85	7.79	1.95	7.35	1.57	7.85	1.57
Time from diagnosis to complete data	3.93	8.99	3.36	8.46	3.67	8.30	2.90	7.13	3.90	9.64	3.45	9.68
Bottom 99% of weights												
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Stabilised Weight value	0.93	3.94E-01	0.87	9.22E-01	0.94	4.32E-01	0.85	9.66E-01	0.94	4.07E-01	0.87	9.34E-01
Previous HbA1c	6.45	0.83	8.15	1.56	6.45	0.83	8.15	1.56	6.45	0.84	8.16	1.56
Previous BMI	30.25	5.94	32.62	6.23	30.25	5.94	32.62	6.24	30.25	5.94	32.62	6.24
Age at onset	63.91	11.53	59.64	11.33	63.91	11.53	59.64	11.33	63.91	11.53	59.62	11.35
Time since study entry	31.03	27.78	20.46	23.33	31.06	27.78	20.49	23.32	31.04	27.78	20.49	23.32
Calendar Year of onset	2006	3.76	2006	3.72	2006	3.76	2006	3.72	2006	3.76	2006	3.72
Baseline BMI	30.88	5.97	32.70	6.19	30.88	5.97	32.70	6.19	30.87	5.97	32.70	6.20
Baseline HbA1c	6.55	1.09	7.72	1.74	6.55	1.08	7.73	1.74	6.55	1.09	7.73	1.74
Time from diagnosis to complete data	4.30	10.22	2.95	7.80	4.30	10.22	2.95	7.83	4.30	10.22	2.95	7.78

Table 7.8 Patient characteristics of those with extreme (top 1%) vs non extreme (bottom 99%) of inverse probability of treatment weights.

7.3.2.3 Models for IPCW

Parameter estimates for the denominator and numerator models to estimate the IPCW are given in appendix 15. Briefly, HbA1c was a strong predictor of censoring, particularly for censoring due to medication change. However, the magnitude of the ORs was much smaller than seen for the treatment models. The distribution of the resulting un-stabilised and stabilised IPCWs are shown in Table 7.9.

	MVRS (A)		**Categorical (C)**	
	Un -stabilised	Stabilised	Un -stabilised	Stabilised
Mean	3.6	1.1	2.0	1.0
Standard Deviation	496.9	15.7	86.7	1.5
1st %ile	1.0	0.4	1.0	0.4
5th %ile	1.0	0.7	1.0	0.7
10th %ile	1.0	0.8	1.0	0.8
25th %ile	1.1	0.9	1.1	0.9
50th %ile	1.2	1.0	1.2	1.0
75th %ile	1.5	1.0	1.5	1.0
90th %ile	2.1	1.1	2.1	1.1
95th %ile	2.8	1.3	2.9	1.3
99th %ile	7.3	2.4	7.2	2.3
Minimum	1.0	0.0	1.0	0.0
Maximum	299,198	8,753	59,553	318

Table 7.9 Distribution of un-stabilised and stabilised inverse probability of censoring weights

7.3.2.4 Combined weights

The distribution of the joint stabilised weights, with and without truncation, are summarised in Table 7.10. As with the IPTW, the joint stabilised weights were truncated at 10 and 0.1, and after truncation the means were 1.00 to 2dp for all covariate specifications.

	MVRS (A)			Simplified (B)			Categorical (C)		
	Stabilised	truncated at 1st and 99th %iles	truncated at 0.1 and 10	Stabilised	truncated at 1st and 99th %iles	truncated at 0.1 and 10	Stabilised	truncated at 1st and 99th %iles	truncated at 0.1 and 10
Mean	150885	0.97	1.00	15413	0.98	1.00	4.3	0.97	1.00
SD	79500000	0.88	1.07	7717480	1.00	1.13	832.5	0.86	1.06
1st %ile	0.04	0.04	0.10	0.04	0.04	0.10	0.05	0.05	0.10
5th %ile	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16
10th %ile	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
25th %ile	0.65	0.65	0.65	0.63	0.63	0.63	0.64	0.64	0.64
50th %ile	0.88	0.88	0.88	0.87	0.87	0.87	0.86	0.86	0.86
75th %ile	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.99
90th %ile	1.35	1.35	1.35	1.32	1.32	1.32	1.42	1.42	1.42
95th %ile	2.14	2.14	2.14	2.05	2.05	2.05	2.25	2.25	2.25
99th %ile	6.83	6.83	6.83	8.03	8.03	8.03	6.48	6.48	6.48
Minimum	0.00	0.04	0.10	0.00	0.04	0.10	0.00	0.05	0.10
Maximum	$7.57x10^{1}$	6.83	10	$7.17x10^{9}$	8.03	10	608003	6.48	10

Table 7.10 Overall distribution of Joint weights for IPTW and IPCW, stabilised and truncated at 99th and 1st percentiles, or 0.1 and 10, for three covariate specifications.

7.3.2.5 Outcome models

Table 7.11 presents estimates of the hazard ratios (HRs) and 95% CI for the effect of metformin vs diet on risk of cancer in patients with newly diagnosed diabetes, for the three model specifications. With standard analysis methods and only a basic adjustment for age, gender, smoking, alcohol and time between diagnosis and study entry, the HR for metformin use was 0.94 (0.86-1.02) for model specification A. A more complete baseline adjustment to this model resulted in the HRs moving closer to 1, though the confidence intervals remained wide. Including adjustments for time updated values of covariates made very little difference to the estimated HRs (model 3). The MSMs with IPTW and then IPTW and IPCW also had very similar HRs, consistent with the results from the standard analysis, though moving to the MSM resulted in a small loss of precision. For specifications A and B, the changes between models were small. Model C showed slightly larger changes in estimated HR between the standard models and the MSMs, but overall, all models were consistent with one another, and with no effect of

metformin on risk of cancer. Model 1 C was suggestive of a small protective effect but this was more likely to be either chance, or residual confounding by age since this was less well adjusted for when only categorised.

	Covariate Specification A		Covariate specification B		Covariate Specification C	
	HR	**95% Confidence Interval**	**HR**	**95% Confidence Interval**	**HR**	**95% Confidence Interval**
Model 1 - basic baseline adjustment	0.94	(0.86 , 1.02)	0.94	(0.86 , 1.02)	0.91	(0.84 , 1.00)
Model 2 - Full baseline adjustment	0.97	(0.88 , 1.07)	0.98	(0.89 , 1.07)	0.95	(0.86 , 1.04)
Model 3 - Baseline and time updated adjustment	0.96	(0.87 , 1.06)	0.96	(0.87 , 1.06)	0.94	(0.85 , 1.03)
Model 4 – MSM with IPTW	0.95	(0.82 , 1.10)	0.97	(0.83 , 1.13)	0.99	(0.86 , 1.14)
Model 5 – MSM with IPTW and IPCW	0.97	(0.83 , 1.12)	0.99	(0.85 , 1.15)	1.02	(0.88 , 1.18)

Table 7.11 Hazard ratios (HRs) for metformin vs diet only on risk of all cancer in patients with T2DM.

Estimates from three standard analysis methods (1-3) and two MSMs. One with IPTW only (4) and one with joint IPTW and IPCW (5).**Model 1 – Minimal adjustment for confounding**: adjustment for age, gender, smoking status and alcohol status and year of onset of diabetes. **Model 2 – Full adjustment for baseline covariates**: Model 1 + baseline adjustment for: HbA1c, BMI, use of other medications in previous year (NSAIDS, statins, antihypertensive drugs), history of chronic kidney disease (CKD) and cardiovascular disease (CVD). **Model 3 – Full adjustment for baseline covariates with time-dependent covariates added:** Models 2 + adjustment for time updated HbA1c, BMI, and history of CVD, CKD and use of other medications in the past 12 months. **Model 4** – As model 2, weighted using IPTW. (MSM with IPTW) **Model 5** – As Model 2, weighted using joint IPTW and IPCW (MSM with IPTW and IPCW). HRs approximated from a pooled logistic regression

7.3.2.6 Sensitivity analysis

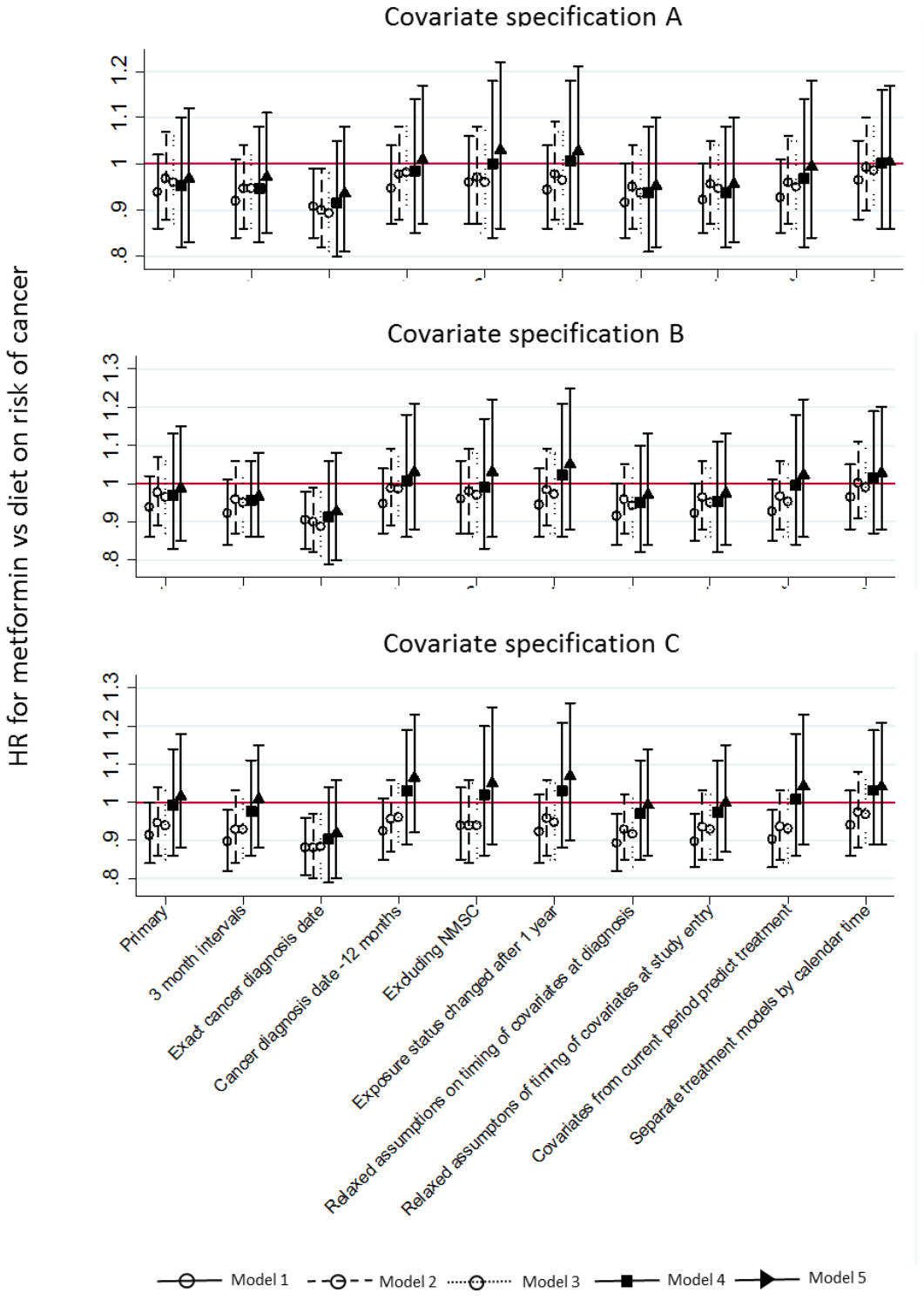
Figure 7.5 provides an overall visual summary of how the sensitivity analyses changed the estimated HR and 95% CI for all five models, for all three covariate specifications. Overall, there were minimal differences to the primary analysis. Full results and confidence intervals for all these analyses are given in appendix 16.

Changing the assumption about how early undiagnosed cancer may affect diabetes severity had very little impact on the result. As a general pattern, using the cancer diagnosis date in the record (0 lag), moved all estimates away from 1 towards a slightly more protective effect, and moving the diagnosis date further forward to 12 months rather than 6 months pushed the results closer to 1 for all models. However, all changes were consistent with random variation and the confidence intervals were all wide.

Estimating IPTW and IPCW separately for different calendar time periods resulted in the need to restrict the analysis to patients diagnosed after the year 2000 only. This was due to small numbers in the previous calendar times meaning the treatment models for some calendar periods had issues with sparsity. There was some suggestion from the results of the models for weights, that the effect of some covariates may be different for differing calendar periods (see appendix 17). For example, the effect of having and HbA1c of 7-8%, 8-10% or >=10% vs <6% post 2005 were all at least 20% higher than for 2000-2005. However, when the weights were applied to the outcome model there was very little difference in the effect of metformin when compared to the primary analysis.

In the remaining sensitivity analyses, changed in estimates from the primary analysis were very small, as shown in Figure 7.5.

Figure 7.5 Estimated HR and 95% CI for metformin vs diet only on risk of any cancer for primary analysis (far left) and 9 sensitivity analyses from MSM with joint IPTW and IPCW



See section 7.2.7.4 for specific details on each sensitivity analysis (order here left to right matches order in which analyses are outlined in section 7.2.7.4) NMSC – non melanoma skin cancer

.

7.3.3 Secondary analyses

7.3.3.1 Cumulative use of metformin

Estimated HRs and 95% CIs for the effect of metformin on risk of cancer over differing exposure lengths are presented in Table 7.12. For all but > 7 years of exposure, the MSM with IPTW (model 4) estimated HRs that were consistent with no effect of metformin on cancer, with decreasing precision as exposure length increased. For > 7 years of exposure, the IPTW model gave some suggestion that metformin was protective against cancer (HR 0.81 (0.44-1.51)) for covariate specification A), albeit with very wide confidence intervals. The standard analyses were generally similar to the MSM with IPTW for all exposure periods. The addition of the IPCW to the MSM (model 5) moved the estimate for > 7 years back towards the null (HR 1.02 (0.50-2.08)) for covariate specification A). All other exposure lengths also had estimates consistent with no effect of metformin on cancer for model 5. All covariate specifications gave broadly similar results.

7.3.3.2 Site specific cancers (prostate, breast, lung, pancreas)

Results of all site specific analyses are presented in Table 7.13. 232 breast cancer events were observed in females with T2DM during the follow up period. A simple baseline adjustment in a standard pooled logistic regression estimated around a 20% decreased risk of breast cancer with metformin use vs diet only (HR 0.81 (0.62-1.06)). Full baseline adjustment made little difference to the estimate. Under model specification C, use of MSMs attenuated the estimate of effect towards the null, while models A and B moved the estimate of effect further away from the null. However, the width of the confidence intervals limits interpretation of any changes that were observed. None of the models investigating the risk of prostate cancer with metformin use in males with T2DM were suggestive of either an increase or decrease in risk. The differences between the model specifications or levels of adjustment, and between standard and MSM analyses were small. For lung cancer, all analyses using standard statistical methods were consistent with no effect of metformin on risk of cancer. Although the confidence intervals are much wider, and therefore interpretation is limited, the weighted analyses appeared to increase the risk estimates to a moderate increased risk of cancer with metformin use.

Length of time on metformin	Model 1 - basic baseline adjustment	Model 2 - Full baseline adjustment	Model 3 - Baseline and time updated adjustment	Model 4 – MSM with IPTW	Model 5 – MSM with IPTW and IPCW
	Covariate Specification A				
0-6 months	0.86 (0.71 , 1.03)	0.88 (0.73 , 1.06)	0.88 (0.72 , 1.07)	0.9 (0.67 , 1.23)	0.93 (0.68 , 1.26)
6-12 months	0.99 (0.83 , 1.19)	1.03 (0.85 , 1.24)	1.02 (0.84 , 1.23)	0.87 (0.68 , 1.12)	1.00 (0.75 , 1.34)
1-2 years	0.92 (0.79 , 1.07)	0.95 (0.81 , 1.11)	0.93 (0.79 , 1.10)	0.95 (0.73 , 1.23)	1.02 (0.78 , 1.32)
2-5 years	0.97 (0.86 , 1.10)	1.00 (0.88 , 1.15)	0.99 (0.86 , 1.14)	1.00 (0.8 , 1.25)	1.09 (0.87 , 1.37)
5-7 years	0.97 (0.77 , 1.23)	1.02 (0.80 , 1.30)	1.01 (0.79 , 1.29)	1.15 (0.79 , 1.68)	0.96 (0.63 , 1.47)
>7 years	0.88 (0.61 , 1.25)	0.94 (0.65 , 1.36)	0.93 (0.64 , 1.35)	0.81 (0.44 , 1.51)	1.02 (0.50 , 2.08)
	Covariate Specification B				
0-6 months	0.86 (0.71 , 1.03)	0.89 (0.73 , 1.07)	0.88 (0.72 , 1.07)	0.92 (0.68 , 1.24)	0.93 (0.69 , 1.25)
6-12 months	0.99 (0.83 , 1.19)	1.03 (0.86 , 1.24)	1.02 (0.84 , 1.23)	0.89 (0.69 , 1.15)	0.89 (0.69 , 1.15)
1-2 years	0.92 (0.79 , 1.07)	0.95 (0.81 , 1.12)	0.94 (0.8 , 1.10)	0.98 (0.75 , 1.30)	1.00 (0.76 , 1.32)
2-5 years	0.97 (0.86 , 1.10)	1.02 (0.89 , 1.16)	1.00 (0.87 , 1.15)	1.01 (0.80 , 1.26)	1.04 (0.83 , 1.31)
5-7 years	0.97 (0.77 , 1.23)	1.03 (0.81 , 1.31)	1.02 (0.80 , 1.29)	1.13 (0.78 , 1.65)	1.06 (0.69 , 1.64)
>7 years	0.88 (0.61 , 1.25)	0.93 (0.64 , 1.34)	0.91 (0.63 , 1.32)	0.87 (0.44 , 1.70)	1.06 (0.51 , 2.18)
	Covariate Specification C				
0-6 months	0.83 (0.69 , 1.00)	0.86 (0.71 , 1.04)	0.86 (0.71 , 1.05)	0.93 (0.68 , 1.27)	0.91 (0.67 , 1.23)
6-12 months	0.97 (0.81 , 1.16)	1.00 (0.83 , 1.21)	0.99 (0.82 , 1.19)	1.00 (0.75 , 1.35)	0.87 (0.67 , 1.12)
1-2 years	0.89 (0.77 , 1.04)	0.92 (0.79 , 1.08)	0.91 (0.78 , 1.07)	0.98 (0.76 , 1.26)	0.96 (0.74 , 1.24)
2-5 years	0.95 (0.83 , 1.07)	0.98 (0.86 , 1.12)	0.97 (0.85 , 1.11)	1.05 (0.84 , 1.31)	1.04 (0.82 , 1.31)
5-7 years	0.95 (0.75 , 1.20)	1.00 (0.78 , 1.27)	0.99 (0.77 , 1.26)	1.06 (0.73 , 1.53)	1.09 (0.70 , 1.68)
>7 years	0.85 (0.60 , 1.22)	0.90 (0.62 , 1.29)	0.88 (0.61 , 1.28)	0.75 (0.41 , 1.36)	1.00 (0.49 , 2.03)

Table 7.12 Hazard Ratios and 95% Confidence interval for risk of metformin use on risk of cancer, estimated by time since first metformin prescription.

	Covariate Specification A		Covariate specification B		Covariate Specification C	
	HR	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**
Breast cancer						
Model 1 - basic baseline adjustment	0.81	(0.62 , 1.06)	0.81	(0.61 , 1.06)	0.82	(0.62 , 1.07)
Model 2 - Full baseline adjustment	0.84	(0.62 , 1.14)	0.85	(0.63 , 1.14)	0.82	(0.61 , 1.11)
Model 3 - Baseline and time updated adjustment	0.89	(0.65 , 1.22)	0.88	(0.65 , 1.20)	0.86	(0.63 , 1.17)
Model 4 – MSM with IPTW	0.80	(0.51 , 1.27)	0.74	(0.47 , 1.18)	0.95	(0.63 , 1.44)
Model 5 – MSM with IPTW and IPCW	0.77	(0.47 , 1.24)	0.73	(0.45 , 1.18)	0.94	(0.62 , 1.43)
Prostate cancer						
Model 1 - basic baseline adjustment	1.07	(0.82 , 1.40)	1.07	(0.82 , 1.40)	1.05	(0.8 , 1.38)
Model 2 - Full baseline adjustment	1.13	(0.84 , 1.53)	1.13	(0.83 , 1.53)	1.08	(0.8 , 1.47)
Model 3 - Baseline and time updated adjustment	1.16	(0.85 , 1.57)	1.13	(0.83 , 1.54)	1.10	(0.81 , 1.51)
Model 4 – MSM with IPTW	1.02	(0.70 , 1.48)	1.01	(0.70 , 1.44)	1.05	(0.70 , 1.58)
Model 5 – MSM with IPTW and IPCW	1.06	(0.73 , 1.55)	1.06	(0.73 , 1.53)	1.09	(0.72 , 1.65)
Lung Cancer						
Model 1 - basic baseline adjustment	0.97	(0.71 , 1.33)	0.97	(0.71 , 1.34)	0.93	(0.68 , 1.28)
Model 2 - Full baseline adjustment	0.99	(0.71 , 1.38)	1.02	(0.73 , 1.43)	0.99	(0.71 , 1.37)
Model 3 - Baseline and time updated adjustment	1.01	(0.72 , 1.40)	1.04	(0.74 , 1.45)	1.01	(0.72 , 1.40)
Model 4 – MSM with IPTW	1.34	(0.80 , 2.26)	1.54	(0.82 , 2.88)	1.24	(0.74 , 2.08)
Model 5 – MSM with IPTW and IPCW	1.45	(0.88 , 2.38)	1.57	(0.89 , 2.79)	1.26	(0.77 , 2.06)
Pancreatic Cancer						
Model 1 - basic baseline adjustment	2.37	(1.31 , 4.29)	2.31	(1.29 , 4.14)	2.25	(1.26 , 4.04)
Model 2 - Full baseline adjustment	2.06	(0.99 , 4.30)	1.95	(0.93 , 4.09)	1.96	(0.96 , 4.03)
Model 3 - Baseline and time updated adjustment	1.61	(0.80 , 3.23)	1.51	(0.76 , 3.02)	1.66	(0.85 , 3.24)
Model 4 – MSM with IPTW	1.92	(0.84 , 4.41)	2.03	(0.88 , 4.67)	2.66	(1.14 , 6.20)
Model 5 – MSM with IPTW and IPCW	2.12	(0.96 , 4.68)	2.39	(1.05 , 5.44)	3.11	(1.24 , 7.76)

Table 7.13 Hazard ratios (HRs) for metformin vs diet only on risk of Breast, prostate, lung and pancreatic cancer.

Estimates from three standard analysis methods (1-3) and two MSMs. One with IPTW only (4) and one with joint IPTW and IPCW (5). HRs approximated from a pooled logistic regression. Breast cancer - females only. Prostate cancer – males only. Due to no events occurring in some subsets of the population, the analyses of lung and pancreatic cancer were restricted to patients over the age of 45 at diagnosis of diabetes who were diagnosed after the year 1995.

For pancreatic cancer, only 50 events were observed during follow up. A simple baseline adjustment estimated an HR of 2.37 (1.31-4.29) for metformin use vs diet, which was expected due to strong confounding by disease severity. A full baseline adjustment moved this estimate back slightly lower (HR 2.06 (0.99 – 4.3 for model specification A). The use of MSMs had little effect on model specification using continuous forms, however for model specification C, the weighting increased the estimated HR to 2.66 (1.14 – 6.20) for IPTW and 3.11 (1.24-7.76) for IPTW and IPCW. However, with the other site specific analyses, the increased width of the confidence intervals in the MSM made the changes between standard methods and the MSMs difficult to interpret. Differences between model specifications could be a result of residual confounding by a less accurate adjustment for HbA1c in the weighting, which is supported by the observation that the differences between model specifications are much smaller for model one which does not include HbA1c.

7.3.3.3 Sulfonylurea monotherapy vs diet

962 patients alive, cancer free and otherwise eligible for study entry initiated a sulfonylurea at time of diabetes diagnosis. A further 3,256 patients initiated a sulfonylurea as a first line therapy at some point during follow up. Total person years of follow up exposed to sulfonylurea monotherapy was 8,757 person years.

Comparisons between distributions of baseline covariate values between patients initiating sulfonylureas at time of study entry compared to diet only (analogous to table 4.3) are presented in appendix 18, table 18.1. The mean HbA1c of patients initiating sulfonylureas was much higher than those not starting any therapy, and the mean BMI was much lower. Post 2005, a smaller proportion of patients were prescribed sulfonylureas compared to before 2005, reflective of the increase in metformin prescribing. The distributions of most baseline variables for those initiating sulfonylureas at study entry were similar to those observed in baseline metformin initiators. The exception was that statins and anti-hypertensive drugs appeared to be used less frequently in patients initiating a sulfonylurea compared to both metformin and diet only.

For many continuous covariates, the associations estimated by the cubic splines were similar to those for metformin. However, the shape of the associations with previous BMI and age at diagnosis were estimated to be quite different (see appendix 18 figures 18.1 and 18.2). The stabilised IPTW had smaller extremes than those for the metformin analysis. The distribution of

the joint censoring weights was similar that of the joint weights in the primary analysis (see appendix 18 tables 18.2 and 18.3).

With just a basic baseline adjustment using standard methods, the estimated HR for the effect of sulfonylurea use vs diet only on risk of cancer was 0.79 (0.67-0.96) for model specification A, and almost identical for B and C. Full baseline adjustment attenuated this protective effect to around a 15% reduced risk for sulfonylurea users compared to diet. In all model specifications, the MSMs appeared to further attenuate the effect, particularly when using joint IPTW and IPCW (Table 7.14). Although the changes between standard methods and MSMs appeared slightly larger than in the metformin analysis, once again the confidence intervals for all models overlapped to a considerable extent.

	Covariate Specification A		**Covariate specification B**		**Covariate Specification C**	
	HR	**95% Confidence Interval**	**HR**	**95% Confidence Interval**	**HR**	**95% Confidence Interval**
Model 1 - basic baseline adjustment	0.79	(0.64 , 0.96)	0.79	(0.65 , 0.96)	0.79	(0.64 , 0.97)
Model 2 - Full baseline adjustment	0.84	(0.68 , 1.04)	0.84	(0.67 , 1.04)	0.82	(0.66 , 1.02)
Model 3 - Baseline and time updated adjustment	0.85	(0.68 , 1.06)	0.83	(0.66 , 1.03)	0.83	(0.67 , 1.03)
Model 4 – MSM with IPTW	0.90	(0.67 , 1.21)	0.89	(0.66 , 1.19)	0.85	(0.64 , 1.14)
Model 5 – MSM with IPTW and IPCW	0.92	(0.65 , 1.30)	0.94	(0.66 , 1.34)	0.89	(0.63 , 1.26)

Table 7.14 Hazard ratios (HRs) for sulfonylureas vs diet only on risk of any cancer in patients with T2DM

HRs approximated from a pooled logistic regression. Estimates from three standard analysis methods (1-3) and two MSMs. One with IPTW only (4) and one with joint IPTW and IPCW (5). **Model 1 – Minimal adjustment for confounding**: adjustment for age, gender, smoking status and alcohol status and year of onset of diabetes. **Model 2 – Full adjustment for baseline covariates**: Model 1 + baseline adjustment for: HbA1c, BMI, use of other medications in previous year (NSAIDS, statins, antihypertensive drugs), history of chronic kidney disease (CKD) and cardiovascular disease (CVD). **Model 3 – Full adjustment for baseline covariates with time-dependent covariates added:** Models 2 + adjustment for time updated HbA1c, BMI, and history of CVD, CKD and use of other medications in the past 12 months. **Model 4** – As model 2, weighted using IPTW. (MSM with IPTW) **Model 5** – As Model 2, weighted using joint IPTW and IPCW (MSM with IPTW and IPCW)

7.4 DISCUSSION

This study aimed to investigate the causal association between metformin use and cancer risk in patients with type 2 diabetes, using routinely collected primary care records. By modelling medication with metformin as a time-varying exposure, and using MSMs with IPTW, it was possible to compare metformin users with patients managing their diabetes with diet alone, while adjusting for time-dependent confounders affected by prior treatment.

Performing the analysis as such, no evidence of an association between metformin use and risk of cancer in patients with type 2 diabetes was found, with HRs for the association close to 1 for all models fitted. This finding was broadly consistent across a range of sensitivity analyses, and also when looking at cumulative metformin use.

The differences between the MSMs and standard analysis methods were minimal in the primary analysis and across a range of sensitivity analyses. Slightly larger differences were observed when examining site specific cancers, though small numbers of events resulted in extremely wide confidence intervals making differences between models difficult to interpret.

The effect of sulfonylureas vs diet on risk of cancer was also estimated in a secondary analysis, and here, although lacking precision, there was a tendency for the results of the MSMs to be closer to the null than the standard analysis methods.

7.4.1 Comparison to other studies

To date, no existing studies that implemented MSMs with IPTW to examine the association between metformin and cancer risk could be found. As discussed in the literature review, existing studies with lowest risk of bias tended to be studies comparing metformin to an alternative first line diabetes therapy such as a sulfonylurea, with adjustment for potential differences in disease severity that may affect both cancer risk and choice of medication at the time of first exposure. Although answering a slightly different question, the results of these studies are consistent with the analysis presented here, in that they found no evidence of an association between metformin use and cancer risk. For example, Van staa et al (2012) [83] found no protective effect of metformin compared to sulfonylureas apart from in the first year of treatment, which is likely to be a result of undiagnosed cancer causing changes in disease severity that result in sulfonylureas being used over metformin. Additionally, a systematic review and meta-analysis of safety data from randomised controlled trials [203] found a

summary RR for metformin use on risk of cancer of 1.36 (0.74-2.49) for trials of usual care/placebo comparators, and 1.02 (0.82-1.26) for all trials (i.e. a mixture of placebo controlled, usual care, and active comparator trials). These results are both consistent with the results from this study. Though it has a wide confidence interval, the RR for metformin vs placebo/usual care is slightly higher than we observed. This difference in estimates could be due to differences between the comparator treatments. Our comparator was no record of medication, which was assumed to be diet only. However, the review pooled placebo and "usual care", where usual care was not clearly defined. Additionally, the review did not limit to trials in patients with diabetes, and also included trials in patients "at risk" of diabetes. The mean follow up of the 13 trials included was 3.58 years, which is also similar to our study.

Other causal methods have previously been used to attempt to better adjust for baseline confounding, and informative loss to follow up. A recent study by Ko et al [93] used inverse probability weighting based on a baseline propensity score (PS) to compare incident metformin and sulfonylurea users for risk of endometrial cancer over a median follow up time of 1.2 years (IQR 0.4-2.3 years). This study did not restrict to patients with diabetes, but in a secondary analysis the authors estimated a HR or 0.88 (0.69-1.16) for diabetic patients without a diagnosis of polycystic ovary syndrome, which is the closest subgroup to the population in this analysis. Their result is broadly consistent with the results presented here, and is most similar to the model with only the basic baseline adjustment (0.93 (0.85-1.01)), which could be explained by the fact that this study did not have information on HbA1c or BMI to include in their propensity score. The authors did not report the distribution of the IPTW in order to compare with what was observed here, however the lack of HbA1c in their models would imply they most likely had less extreme weights. They did report that progressive trimming of the PS distribution made very little difference to the overall estimate.

In another study, new users of metformin were compared with new users of sulfonylureas, where like the current study, patients were censored if they initiated any other treatment. IPCW were used to adjust for loss to follow up [28] but the weights only calculated at time points of 1, 3 and 5 years. Also based on patients with early T2DM in the CPRD, this study serves as a good comparison to our analysis. The authors estimated the risk for cancer for metformin vs sulfonylureas to be 0.96 (0.89-1.04) without IPCW, and 0.94 (0.85-1.04) with IPCW (for censoring by 5 years of follow up). Although they compare metformin with sulfonylureas, these results are similar to the findings from this analysis, and also consistent in that they also observed few differences between the models with IPTW alone, and IPTW with IPCW.

Although these studies are consistent with one another and with our results, all these studies also had relatively short average follow up within which a true effect of metformin may have been observed, and it is possible that this explains the estimated null effects.

7.4.2 Comparing MSMs to standard analysis methods

The results from the models using MSMs with IPTW and also joint IPTW and IPCW weights, were compared with results using standard analysis methods that in theory would not correctly adjust for time-dependent confounders affected by prior treatment. There was an a priori belief based on existing literature that covariates such as BMI, HbA1c, and other measures of diabetes severity satisfied the definition of time-dependent confounders affected by prior treatment. In addition, initial analyses (although univariate) suggested that there were associations between these covariates and cancer in our data (7.3.1). However, the MSMs produced results that were similar to those obtained via standard analysis methods. The standard pooled logistic regression model with time updated exposure and full baseline adjustment estimated an HR for metformin use of 0.96 (0.87 – 1.06) for covariate specification A, compared to 0.97 (0.84 – 1.13) for the MSM with joint IPTW and IPCW weights. None of the varying levels of adjustment for confounding via standard analysis methods produced noticeably different results.

An obvious reason for the lack of difference is that the post baseline confounding between initiation of metformin and cancer incidence is not as strong as initially hypothesised. One reason for this could be that not enough patients were initiating treatment far enough away from baseline for the values of the confounders to change sufficiently to make post-baseline confounding apparent. The analysis in chapter 6 suggested that the median time to treatment after study entry was 2 months. Considering HbA1c is a measure of long term glucose control over approximately 3 months, this means that the post baseline confounding would at most affect around 50% of the population. In the analysis of cumulative medication, the differences in estimates for > 7 years exposure between standard methods and MSM with both IPTW and IPCW were more noticeable, which supports the idea that longer follow up may be needed for the time-dependent confounding to become apparent.

HbA1c was by far the strongest predictor of treatment initiation through time, which is unsurprising due to the UK diabetes treatment guidelines being primarily based on observed HbA1c levels. However, for there to be significant confounding, there must be clear association between HbA1c and risk of cancer. Review articles examining diabetes and cancer risk overall

[19, 20] suggest that the epidemiological evidence for hyperglycaemia and cancer risk indicates an association, but not necessarily causality. They suggest for example, that in non-insulin deficient situations, hyperglycaemia is a proxy indicator for hyperinsulemia, which is a more plausible causal risk factor. However, they also acknowledge that many cancers require glucose for energy, so a causal association should not be completely disregarded. This suggests that HbA1c should certainly be considered as a confounder, and therefore requires the use of MSMs to model time updated exposure. In a practical sense however, the necessity of the adjustment (in both the MSM and standard analysis) then depends on the strength of the association. Two meta-analyses, one in clinical trial data and another using observational research provide possible quantifications of the effect of hyperglycaemia on cancer risk. A meta-analysis of safety data from clinical trials in patients with T2DM comparing intensive vs normal glucose control were suggestive of a marginal decrease in risk of cancer incidence with tighter glucose control, albeit with a relatively wide confidence interval, pooled HR (0.91, 0.79-1.05) [204]. In overweight patients with T2DM, there was larger decrease in risk of cancer mortality with intensive glucose control (HR 0.74, 0.37-1.48). Although the authors are cautious and overall suggest there is no evidence of an increased risk of cancer with poor glucose control, their estimates do not exclude it. A larger meta-analysis of 14 studies looking at various site specific analyses reported associations between raised HbA1c and increased risk of cancer that were much larger in magnitude [205]. However, the quality of the included studies has not been thoroughly examined and the same issues of time-dependent confounding may also be relevant to these studies. Based on these findings, it is certainly clear that HbA1c should be considered a potential time-dependent confounder, however for the MSM to be beneficial over standard methods, the association must be apparent in the data. In the data used for this analysis, there was some evidence of a univariate association between current HbA1c and risk of cancer, though the increase in risk was relatively consistent around a 10% increase for all HbA1c categories vs 6% (Table 7.4). Overall, if the effect of raised HbA1c on cancer risk is only around a 10% increased risk, it is quite possible that when taking into consideration factors such as measurement error and frequency of HbA1c measurements in CPRD, the effect of adjusting for HbA1c as a time-dependent confounder via use of MSMs may not have a prominent effect on the overall estimate of risk. However, with the current literature providing wide-ranging estimates for the actual association, it was necessary and useful to explore the effect that use of MSMs may have had on the estimated risk.

Similarly, BMI has been shown to be associated with risk of cancer [195, 196, 206] . Therefore, this was an important potential time-dependent confounder. Original guidelines suggested that

metformin should be prescribed to more overweight patients, due to its weight reduction properties, and in terms of the secondary analysis, that sulfonylureas should be prescribed more cautiously in overweight patients due to the potential for it causing further weight gain. However, only about 4% of the population studied were diagnosed with T2DM before the year 2000, beyond which BMI was less clearly associated with choice of treatment. Therefore, although it can still be useful to adjust for in the model, it may have made less of a difference between the standard methods and MSMs than initially hypothesised. Another issue may be that the association between BMI and risk of cancer differs for different cancer types [195, 196]. By combining all cancers into a single outcome, differing effects of BMI on different cancers combine in ways that are hard to predict. This could have reduced additional benefit of adjustment for confounding by BMI (both baseline and time dependent). The slightly larger observed changes between standard models and MSMs in the site specific analysis support this possibility, though the imprecision of these analyses limits interpretation of these differences.

7.4.3 Validity of assumptions

The discussion below focuses on the assumptions needed for valid interpretation of the results from the MSMs, providing further insight into the potential reasons for lack of differences observed between the MSM and the standard analysis methods.

7.4.3.1 Unmeasured confounding.

Importantly, there may be key confounders missing from the model. Although discussions with experts concluded that the most important measured confounders in CPRD had been considered, there are potentially some confounders that were not available in CPRD that could have had an impact, such as levels of physical activity and diet. These factors may affect risk of cancer independently of BMI [207-210], and be associated with treatment through time, due to their influence on the GP's perception of whether a patient would be likely to adhere to a diet and lifestyle intervention.

Residual confounding when using primary care records is likely for factors such as smoking and alcohol despite adjustment, due to a lack of detailed information on quantity (i.e. amount of smoking/alcohol). For example, although groups of current, ex or never smokers has been shown to be relatively well recorded in primary care [211], the smoking quantity is rarely

recorded, and when it is, it is likely to be under-estimated due to self report [212]. Similarly, although alcohol quantity is better recorded, it may also be subject to social desirability bias. This lack of detail is likely to be an important limitation when looking at cancer, particularly in terms of smoking. In particular, since smoking is associated with incidence of type 2 diabetes [213], the prevalence of smoking is high in patients with T2DM. In this study, only 40% of patients were never smokers at the time of diabetes diagnosis, which is around 10-15% lower than the proportions reported by the ONS for similar age ranges between 2000 and 2016 [214]. Therefore, the residual confounding due to limited information on smoking quantity and duration in ex and current smokers may affect this analysis more than an analysis within the general population. To investigate how much residual confounding may have been present, it may have been useful to re-run the analysis looking at non-smokers only. However this would have reduced the numbers available to less than 50% of the original sample and would have had limited power. This would also be problematic if the effect of metformin on cancer risk was modified by smoking, as the marginal effect of metformin would then be different to the effect in non-smokers alone, making the results incomparable. Further to the issue with lack of data on duration and quantity, which is a limitation of the CPRD, in this study, smoking and alcohol were considered as fixed baseline confounders only. Although all past information from baseline was used to determine the most appropriate category, this was not allowed to change after study entry, and so patients quitting smoking after diabetes diagnosis would not have been captured. Although given the length of follow up, it is unlikely that any changes in smoking or alcohol after study entry could have substantially modified their cancer risk, there is the possibility of further residual time-dependent confounding from this. However, it is unlikely that a lack of time updated information on smoking and alcohol alone would completely explain the lack of difference between the standard methods and MSMs.

Issues of residual confounding apply to other covariates such as use of statins, anti-hypertensives and NSAIDS, where dosage was not taken into account. In addition, the possibility of misclassifying patients in terms of CKD and CVD cannot be excluded.

7.4.3.2 Misspecification of the weighting models

It is also possible that the variables specified in the model had inappropriate forms/parameter specifications. Three alternative specifications were used to fit the weighting models, and the estimates of risk of cancer with metformin use were very similar for all. This suggests either that the model was in fact relatively robust to different model specifications, and as such that this is

not the reason for the lack of difference from the standard methods, or that none of the specifications were appropriate. However, one of the specifications used a data driven approach to fit cubic splines, and therefore should have modelled the association between the covariates and treatment extremely finely. This specification did result in some very large weights, even after stabilisation, and the truncation of these weights may have attenuated the added benefit of the flexible form for the weighting model. Having said this, the truncation used in the primary model was to truncate at 10, which was beyond the 99th percentile, so is unlikely to have substantially removed any effect of weighting from the parameter estimate.

7.4.3.3 Misspecification of the MSM (outcome model)

As explained in section 4.4.2.5, it is particularly important that the outcome model is correctly specified in terms of the effect of time, the baseline confounders and treatment. The three different covariate specifications were also used for the outcome models to look at how sensitive the results may be to potential misspecifications. The same spline functions that were used in the treatment models were applied to the outcome models. It is acknowledged that this would not be appropriate if the spline function was too simple to model the complex association between the covariate and risk of cancer, however this was not considered to be an issue here, as checks indicated that using more flexible splines did not estimate very different shapes of association for any of the covariates in the outcome model (appendix 14). The use of the same simplified continuous parameterisations in specification B may have been more problematic, particularly for baseline BMI that was simplified to a linear term. However, the estimates of effect from all outcome models, including the categorical specification were very similar. As such, it is unlikely that any major bias was introduced from incorrect specifications for some covariates in any one of these model specifications. In terms of how treatment was specified, the primary analysis assumed a constant effect of treatment in line with most of the existing studies that have examined the same association. The potential misspecification of this was investigated by looking at cumulative use of metformin in a secondary analysis, but this did not show any strong suggestions of differing effects through time. Finally, no interactions were included in the outcome model due to the large number of possible combinations of covariates, lack of power, and because there were no clear a priori hypothesised interactions. Although unlikely to be a major limitation, it is acknowledged that this could have resulted in some misspecification of the outcome model.

7.4.3.4 Positivity

Initial descriptive statistics suggested that HbA1c may cause issues with near violations of positivity, as high HbA1c is such a strong indicator for being treated with metformin. As expected, even after stabilisation there were extremely large weights for some individuals, usually driven by the characteristic of having a high HbA1c but not being treated with metformin, or having a very low HbA1c but initiating metformin. However, just a slight truncation of the weights reduced the mean of the stabilised weight to 1, and therefore it was not considered necessary to attempt alternative approaches to reduce the extreme weights. Another approach to address the near violations of positivity could have been to trim the actual study population according to pre-specified HbA1c criteria, and so remove the patients that had the most extreme weights. This would have resulted in a population whose HbA1c levels remained within a stricter range and as such would have changed the generalisability and interpretation of the estimates of effect, though an advantage over weight truncation is that this change in interpretation can be clearly defined. It would however, have been difficult since HbA1c is a time-dependent measure, and as such a decision would have been needed as to whether the patient should be censored when their HbA1c goes outside the range, or whether to remove a patient if their HbA1c ever went outside the range. The first approach would result in needing another level of censoring in the IPCW models, which would not only increase computational intensity but also mean that censoring would become highly dependent upon HbA1c, and therefore would likely increase the size of the IPCW. The second approach would also be problematic as it would result in excluding patients based on future information. The coarser form of HbA1c used in covariate specification C did reduce the size of the extreme weights, but not enough to avoid the need for truncation, and so in this case altering the model specification had only a minimal effect on improving positivity. Overall, the process of fitting treatment and censoring models requires balance between having an adequately specified model and one that does not result in violations of positivity. Considering the robustness of the models to different covariate specifications, and the need to only truncate the most extreme 0.6% of the joint weights, it seems that overall this was sufficiently achieved.

7.4.4 Visit frequency

A key issue with using data from primary care, is that the probability of treatment can only be non zero if the patient visits their GP. As discussed in chapter 6, HbA1c and BMI were only

recorded every 6-8 months; and this was less frequent in patients who were not initiating metformin, which could suggest that the lack of treatment was due to non-attendance. If the reason for not visiting impacts the risk of cancer, then further bias will be introduced. Since patients are more likely to visit their GP if they are generally unwell, have particular concerns, or think they may developing cancer, it is possible this could have been the case in the present study. However, the total number of GP consultations in the past year was similar between those initiating metformin and those remaining untreated, which argues against attendance patterns being a major determinant of treatment initiation. Also Since 2005, the quality outcomes framework [215] has incentivised regular visits for patients with T2DM, meaning that all such patients should have more regular visits. A more formal assessment of this issue could be achieved by extending the causal methodology used here to model visit frequency in addition to probability of treatment and probability of censoring, in a similar way to how the joint effect of two treatments are modelled in other settings [216]. This would require consideration of whether it is important to differentiate between complete non-attendance vs having intermittent missing data on specific covariates.

7.4.5 Interval data format

It was necessary to split the time into discrete time intervals in order to estimate time updated IPTW and IPCW. This required decisions to be made on length of interval, how to assign measurements to the intervals, and how to deal with intervals in which multiple events occurred. For this study, 1 month intervals were used for the primary analysis. This choice was a balance between a plausible visit frequency (likely to be longer than one month) and an interval length that would ensure variables measured in the previous interval could plausibly affect treatment and risk of outcome in the following interval. Sensitivity analyses using wider intervals of 3 months made little difference to the estimates of effect of metformin on cancer risk from the MSMs, suggesting that in this case, the interval choice was not important.

Related to this, another sensitivity analysis assessed the difference between using data from the previous interval to predict treatment and censoring (primary analysis), or data from the same interval (sensitivity analysis). This tested how important it was to ensure strict temporality between covariate measurement and treatment initiation. Again, there was little difference to the primary analysis, suggesting that the data in the previous interval was sufficient to capture risk of outcome between those initiating or not initiating treatment in the following interval. Having said this, the lack of difference between the sensitivity analysis and primary analysis

could also just reflect the lack of updating of key covariates in every interval. As long as it can be assumed that any large changes in covariates that will have a strong impact on treatment decisions and risk of cancer are likely to be recorded, then the approach of LOCF is tolerable. Via both initial investigations of the data (as presented in 6.3.2) and discussions with GPs and other clinicians, it was decided that this was a reasonable assumption. Having said this, the possibility that infrequently updated data on HbA1c and BMI has caused bias in our estimates cannot be excluded. This could have been investigated further by obtaining a sub-sample of patients with more regular measurements, to see if the weighting models were substantially different. The reason this was not done was because patients in such a sub-sample are likely to have different characteristics that cause them to have measurement taken more frequently, and so the results of such a comparison would be hard to interpret.

A final discussion point related to interval set up was how to deal with censoring and cancer events occurring in the same interval. With EHR data, unless the censoring event is death or transfer out, we do not actually lose the patient to follow up, and it may be that a patient has the cancer event either before or after censoring but in the same interval. If this occurred, the decision was made to censor the subject and not count the cancer as an event. This was done because formally, the MSM evaluates the risk of cancer conditional on remaining uncensored to the end of the discrete interval. However, if censoring events such as medication change are a result of the cancer diagnosis, then this censoring rule may result in some cancer events being missed. Further, if the available covariates do not adequately balance this potential informative censoring then there may be some bias introduced into the estimated effects of treatment. In the primary analysis, because diagnosis dates were brought forward by 6 months, the applied censoring rules were not considered to be a major limitation. However, outside the scope of this study, if the research question examines more acute outcomes, the ordering of censoring and outcomes within an interval may be important to consider (and will be considered in later chapters).

7.4.6 Other Limitations

7.4.6.1 Missing data

In an ideal clinical trial, all patients would be followed up from time of diabetes diagnosis. However, to correctly balance the risk of outcome between those initiating and not initiating metformin, data are needed on all relevant covariates from the beginning of the study. In our

data, since not all patients had complete data at the time of their diabetes diagnosis, some patients had what could be thought of as “delayed entry” into the study. This was adjusted for with a variable representing time between diagnosis and study entry. Using this approach compared to using only patients with complete data at the time of diagnosis increased the sample size by about 20,000 patients, but there are potential limitations. All patients that initiate treatment before they have complete data must be excluded, since the covariates at study entry may have already been affected by treatment. This could induce selection bias if the reason for not having complete data is related to risk of cancer, since exposed patients with lower/higher risk of cancer would be systematically excluded. Considering around 27,000 patients were excluded due to initiating treatment before having complete data, this issue could have had a big impact on the results.

An alternative approach could have been to use multiple imputation to impute baseline covariate values for all subjects and then use LOCF as if they were true values. This however relies on the assumption that the data are missing at random, so that the probability a value is missing is not related to the value itself, after conditioning on other available covariates. In this context, it may be a fair assumption, since one of the main predictors of missing data at time of diagnosis was calendar time, though in general the missing at random assumption is unlikely to be perfectly satisfied in data recorded for clinical care. There is a limited amount of research on the use of multiple imputation with MSMs and how it may impact the estimation of the treatment weights. However, recent research into multiple imputation in a propensity score context has shown the importance of implementing the entire process of calculating the propensity score and IPW within a multiple imputation framework before combining the imputations, to avoid bias [217]. It is likely that the full multi-step process of fitting a marginal structural model would also have to be done within an imputation framework, and the computational cost of this in such a large sample may be prohibitive.

7.4.6.2 Computational limitations

Other issues relating to the size of EHR data sets were also apparent during the model implementation process. With a median follow up time of 3.7 years, each patient had an average of 44 rows of data using one month intervals. This meant an overall data set of more than 2.2 million observations, resulting in computation times of around 30 minutes to calculate IPTW and IPCW weights, and a minimum of 10 minutes to fit the outcome models. Although this was manageable with a high specification PC, for this study, the cohort had relatively strict inclusion

criteria, as it was important to identify patients with incident diabetes with a clearly identifiable date of diagnosis. For EHR studies in less specific populations, for example, in users of statins, sample sizes could be much larger, and using MSMs with IPTW (and IPCW) may be less feasible without some restriction of the study population.

7.4.6.3 Follow up time

As already eluded to in 7.4.1, the average follow up time of patients in this study was relatively short. With a mean time of just under 4 years, there may have been insufficient follow up to detect any causal effect of metformin on cancer. By stratifying by length of exposure to metformin in a secondary analysis, it was possible to obtain an estimate for the effect of 5-7 and > 7 years of metformin use on cancer risk. This analysis did not find any long-term effect, however due to fewer numbers at risk; the precision of these estimates was low. In the systematic review of clinical trials conducted by Stevens [203] only one trial had greater than 10 years follow up, and this study looked at cancer mortality rather than incidence [31].

7.4.6.4 Outcome definition

The potential for misspecification or measurement error for covariates has already been discussed, however the possibility of information bias arising from the definitions of exposure and outcome used in this study also require evaluation.

The definition of cancer diagnoses used was based on an existing algorithm used in previously published work [195]. Although not formally tested for validity, the broad starting inclusion followed by manual sorting of codes with checking performed by two separate researchers including a clinician means it is likely to accurately identify cancer diagnoses with good sensitivity and specificity. The cancer definition is limited however, in that is only uses information from primary care. A previous study by Boggon et al [218] estimated that 92% of cancers (excluding NMSC) recorded in the CPRD were also recorded in at least one of the National Cancer Data Repository (NCDR), Hospital Episode Statistics (HES) or death certificates. This suggests that cancer diagnoses taken from CPRD alone have relatively good concordance with external sources, and have a low false positive rate. This is strengthened by a further systematic review of validation studies in the CPRD found that from 7 studies examining the validity of cancer diagnoses, the proportion of confirmed cases ranged from 74% to 100%

[188]. However, feedback to GPs may be imperfect, if for example, hospital letters to the GP are just scanned and attached to the patient's record, or the diagnosis entered as free text. For example, in the study by Boggon et al, although 94% of cancer records within the cancer registry were recorded in some way in the CPRD, 11% were in free text only, and so our algorithm would not have captured these, meaning we may have underestimated the number of cancer cases overall by around 10%. Further, the study found that concordance between the two data sources was lower for certain cancer sites. For example, for colorectal, lung and pancreatic cancers, there were more cases recorded in the NCDR. Therefore, in some of the site-specific analyses conducted, it is possible that we will have missed a larger proportion of cancer events. However, the effect this is likely to have had on the estimated effects of metformin is small, since the hazard ratio remains unbiased when the misclassification only affects sensitivity and not specificity [219]. Therefore, the main issue with the underestimation of cancer events is the reduced precision. Linkage to cancer registries from the CPRD is only available for about 60% of the patients within the CPRD. If we were to restrict the analysis to patients eligible for linkage, despite getting a more complete capture of cancer outcomes within that population, the overall number of events captured would likely be smaller due to the reduced sample size overall. It was therefore felt that using cancers recorded within the CPRD only was a reasonable approach.

It is also possible that some of the cancer diagnoses in CPRD would be delayed in their entry as there would be a time lag between diagnoses in secondary care, and this diagnosis being communicated to the GP. This however would be partially taken into account by the fact that in the primary analysis, the cancer diagnosis date was brought forward by 6 months (albeit because this was primarily to make sure that changes in disease severity due to un-diagnosed cancer were accounted for). Whether this lag was appropriate is another point for consideration. Sensitivity analysis of no lag and 12-month lag did not show large differences in estimates, though there was some indication that the 12-month lag pulled estimated metformin/cancer associations closer towards 1, and using no lag made metformin appear slightly more protective. A possible explanation for this is that changing the cancer diagnoses date changes the time point at which we compare medication with the rest of the risk set. Patients already on metformin are more likely to be censored for starting a treatment other than metformin than those not on treatment, and without the lag, we are more likely to censor because of treatment before we observe the cancer. Removing the lag on the cancer diagnosis may therefore systematically remove metformin users that go on to get cancer. The more the cancer date is brought forward, the more of these cancers are then included. Although the CIs were wide and the difference

between these analyses were small, the change was consistent with what would occur should pre-existing cancer have an impact on medication intensification, suggesting that use of the lag was warranted. Potentially, larger lags of 2 and 3 years could also have been explored to see if this continued to increase the estimates.

The main definition of all cancer in this analysis included non-melanoma skin cancer. This was done initially to improve power, and because there was no reason to think that any effect of metformin might be different for this specific cancer type. However, the association with the risk factors for cancer may be different or weaker for this cancer type. Removal of this kind of cancer from the outcome did not alter the estimated effect of metformin on cancer risk substantially for any of the models fitted, but there some suggestion of a larger difference between the MSMs and standard methods, particularly for the joint IPTW and IPCW. This is consistent with what would be expected upon the removal of events that have no strong association with the risk factors that also influence treatment decisions, though the differences that emerged could have been due to chance since the magnitude of the changes was very small.

7.4.6.5 Exposure definition

The definition of metformin used in this study was chosen for simplicity and to reduce complications with estimating the IPTW. Specifically, only a single prescription was required to be considered exposed, and then it was assumed that the patient remained exposed until there was evidence of a change in medication. However, an obvious limitation of this approach was that it did not take into account that some patients may not adhere to their prescribed medication. Censoring patients if they switched to any other treatment reduced the issue of misclassifying person time exposed to other medications as person time exposed to metformin, but could not account for the situation in which a patient cannot tolerate metformin and so stops and/or reverts to a diet and lifestyle regime. Other options for defining exposure could have included requiring a minimum number of prescriptions before considering a patient exposed, or looking at cumulative use based on actual number of prescriptions. The problem with the former option in the context of using MSMs, is that it becomes unclear how to estimate the probability of treatment in the period between the first prescription and the prescription that confirms exposure.

Another limitation of the way exposure was defined in this study is that dosage of metformin was not taken into account. Aside from requiring more complex models for IPTW, when using

prescription data, dosage must be estimated using information on number of tablets per prescription, tablet dosage, and time between prescriptions. This would raise the complication of deciding what dosage to apply to each interval, and would have to be implemented carefully to avoid using future information to define dosage level. Although beyond the scope of this thesis, investigations into the effects of dosage could be explored in further research.

7.5 CHAPTER SUMMARY

This chapter has presented an application of MSMs with IPTW to assess whether metformin may be causally associated with cancer risk in patients with T2DM, using data from the CPRD.

The key findings of this analysis were as follows:

1. .All models produced results consistent with no effect of metformin on cancer risk, though due to wide confidence intervals, a small protective effect (or harmful effect) could not be excluded. This finding was consistent across models assuming a constant effect of treatment through time, and also analyses that looked at cumulative use of medication.

2. Only minimal differences in estimates of effect of metformin on cancer risk were observed between the MSMs and standard analyses methods. Possible reasons for this could be unmeasured confounding, residual confounding from poor data quality or misspecified models, or weaker time-dependent confounding than anticipated.

3. Secondary analyses of site specific cancers and of sulfonylurea use instead of metformin yielded larger differences between standard analyses methods and MSMs, but the differences were inconsistent between analyses, and very large confidence intervals made sensible interpretation of any changes difficult.

4. Results were robust to a wide range of sensitivity analyses, though this could be partially due to there being no estimated effect of metformin on cancer, meaning that different modelling decisions that in other situations may impact the causal estimate had no effect here.

This study did not suggest any causal effect of metformin on cancer risk. Despite the acknowledged limitations, this adds weight to the existing evidence that suggests that the large protective effects previously observed were not causal. In the following chapter, further analysis aims to identify whether the apparent lack of difference between the MSM and standard analyses observed here was due to poorly developed weighting models, or because the issue of

time-dependent confounding was not important in this particular epidemiological question. The methodology will be applied to different questions regarding metformin use in T2DM, where stronger time-dependent confounding would be expected. Specifically, the next chapter describes an analysis the effect of metformin on risks of MI, stroke, all-cause mortality, and longitudinal glucose control. Assessment of the use of causal methods for these questions will be further enhanced by using existing evidence from randomised controlled trials for comparison purposes.

8 MSM WITH IPTW TO EXAMINE EFFECT OF METFORMIN AND SULFONYLUREA USE ON MORTALITY, CARDIOVASCULAR ENDPOINTS AND LONG TERM HBA1C CONTROL

8.1 AIMS AND OBJECTIVES

The review of the literature given in chapter 3 determined that despite a number of meta-analyses of both clinical trial data and observational studies, there was no firm conclusion about the efficacy/safety of metformin/sulfonylureas on cardiovascular outcomes and all-cause mortality. Importantly, the number of trials and observational studies comparing metformin or sulfonylureas to a diet and lifestyle intervention or placebo was relatively small. In general, the meta-analyses for trials or observational studies that included the most data were for comparisons of metformin vs sulfonylureas. Due to the potential differing effects of metformin and sulfonylureas, such a comparison is unable to distinguish the individual causal effects of the two medications.

The primary aim of this analysis was to estimate the efficacy of first line therapy with metformin and sulfonylureas separately in comparison to diet, with respect to MI, stroke and all-cause mortality using MSMs with IPTW and IPCW to control for time-dependent confounding affected by prior treatment. Time-dependent confounding was anticipated to be greater for these outcomes, due to the stronger association between diabetes severity and cardiovascular complications and mortality. As a long term randomised trial including the same outcomes, the UKPDS study [31] was chosen to serve as a comparison for this analysis.

The secondary aim was to explore whether a clear benefit of using MSMs in data from CPRD could be demonstrated. To do this, it was necessary to examine a question where time-dependent confounding is highly likely, and the true causal effect more clearly established. As far as is known, such an example using EHR data has not been presented to date. It is well recognised that both metformin and sulfonylureas do lower blood glucose levels [220], and the UKPDS study also examined the effect of metformin and sulfonylureas on long term HbA1c. Here, the outcome itself (HbA1c) is a key time-dependent confounder affected by prior treatment, so a strong a priori assumption exists that the standard methods would be biased. Therefore, in addition to the cardiovascular outcomes and mortality, the effect of metformin and sulfonylureas on long terms HbA1c control was also investigated.

8.2 METHODS

8.2.1 Study population

The study population was the same source population used in the previous chapter, namely a cohort patients with incident T2DM, aged between 30 and 90 years and free of any cancer at the time of diabetes diagnosis. However, for this and the following chapter, the data were updated to incorporate the latest data cut available from CPRD at the time of analysis. Therefore, these chapters use an underlying cohort from the January 2016 extract of CPRD, using the same algorithms previously described (see section 5.2). As before, the small number of patients with missing alcohol and smoking data at baseline were excluded. Study entry (baseline) for all other patients was defined as the end of the first interval after diabetes diagnosis that the patient had complete data on all covariates.

Patients were followed up until the event of interest, death (if not the event), transfer out of practice, initiation of any diabetes medication other than metformin or sulfonylureas, or the last data collection date. As before, last collection date was not included as a censoring event for the IPCW.

Two additional restrictions were placed on the study population for these analyses. Firstly, although in the previous chapter, estimation of weights separately by calendar period did not appreciably change the estimated effect of metformin on cancer, it did suggest that the effect of covariates on treatment may be different by calendar period (see appendix 17). Since time-dependent confounding was expected to be stronger for the outcomes in the present chapter, it was felt that the separate treatment models by time should be used. In order to do this, only patients diagnosed after the year 2000 could be included, due to small numbers in earlier calendar periods causing difficulties with model convergence. Secondly, initial investigations (see appendix 19) suggested that follow up should be restricted to a maximum to 10 years to reduce the risk of severe positivity violations.

8.2.2 Exposure definition

The way in which exposure to metformin and sulfonylureas was defined has been described previously (see 7.2.1 and 7.2.7.5.3). Briefly, the exposure status of an individual could update once at the date of the first metformin or sulfonylurea prescription; from diet to metformin, or

diet to sulfonylureas. As before, further prescriptions were not needed to confirm or maintain exposure status. In contrast to the previous chapter, where patients who had started metformin were censored if they then initiated a sulfonylurea, and vice versa; in this analysis the patient was only censored if they initiated something other than a sulfonylurea or metformin. Use of an intention to treat (ITT) approach based on first line initiation not only increased follow up time to observe the cardiovascular events, but it was also more reflective of the UKPDS study protocol, where patients could intensify with one of the other study drugs if necessary. An "as treated" effect, where patients were censored at any treatment switch (other than their first initiation with either metformin or sulfonylureas) was estimated in a sensitivity analysis.

8.2.3 Outcome definitions

8.2.3.1 MI and stroke

Occurrence of MI and stroke was identified from the patient's primary care record only. These were identified using the Quality and Outcomes Framework (QOF) preferred and alternative reporting codes [198]. The full code lists are presented in appendix 12. Both fatal and non-fatal stroke and MI were included. The diagnosis of stroke was broad, and included both ischaemic and haemorrhagic stroke, and the less serious event of transient Ischaemic attack (TIA). Although not usually included within the definition of stroke, TIA was included to improve power. It was felt that this would be reasonable since it has very similar risk factors to stroke, and is a strong predictor of subsequent stroke [221]. A sensitivity analysis was conducted to look at whether excluding TIA had an impact of the estimated effect of metformin (see 8.2.5.3.) The outcome was defined as the first occurrence of the event after study entry. Patients with a history of the outcome at the time of their diabetes diagnosis were not excluded, but CVD history was adjusted for in the modelling (see 8.2.4). A small number of participants were excluded where the outcome of interest occurred between time of diabetes diagnosis and study entry (51 for MI, 79 for stroke), since these patients would not have still been at risk had they been followed up from time of diagnosis (see section 4.4.2.1).

8.2.3.2 All-cause mortality

Death and date of death was identified using a specific "death date" variable specified within the database. This variable is derived by the CPRD via an algorithm that takes information from

the patient's clinical and administrative records to identify whether the patient has died, and the date that it was entered onto the system. Linkage to ONS was not used to obtain cause of death and as such only all-cause mortality was studied.

8.2.3.3 HbA1c

A variable capturing time updated HbA1c was constructed for each patient, as described in the previous chapter (see section 7.2.3.2.1 and appendix 12). These HbA1c measurements were used as a repeated measures outcome, to model the trajectory of HbA1c through time. Only intervals in which HbA1c was observed were included in the vector of outcomes values for each patient.

8.2.4 Covariates

All covariates from the previous analysis were included in the weighting models, namely age, gender, calendar period of diabetes onset (pre or post 2005), baseline smoking status, baseline alcohol consumption, use in the previous year of anti-hypertensive drugs, statins, NSAIDS (both baseline and time updated); previous history of any CVD or CKD (both baseline and time updated); HbA1c and BMI (both baseline and time updated).This decision was based both on a priori knowledge about potential associations with outcome; and from examining the observed associations with the outcomes of interest (see sections 8.2.5.1 and 8.3.2). Additional covariates were included to further adjust for possible time-dependent confounding, and are described below.

8.2.4.1 Recent history of CVD

In the previous chapter, both baseline and time updated history of CVD included stroke, MI, TIA and ischaemic heart disease. However, for the outcomes of interest in this analysis, it was felt that it may be important to distinguish between history of the outcome in question, and other CVD. Therefore history of MI, stroke (including TIA) and other CVD were entered separately. Variables indicating a recent history, as defined by occurrence of a code in the last 3 months, were also included due to the fact that recent history of an event is likely to substantially increase the risk of another event. As in the previous chapter, these variables were considered

time-dependent. Time updated history of stroke was omitted from the models relating to stroke as an outcome, and time updated MI from the models for MI, since in these cases any update after baseline would represent occurrence of the outcome.

8.2.4.2 Systolic blood pressure (SBP)

Both baseline and time updated SBP were included due to the strong association between cardiovascular events and hypertension. Full details of how SBP was extracted and cleaned are available in appendix 12. As with HbA1c and BMI, the baseline SBP was taken to be the closest measure prior to study entry. For patients entering the study at time of diabetes diagnosis, the baseline SBP was required to have been within 6 months otherwise it was considered missing. Values were truncated at the 1st and 99th percentiles of their baseline distributions to avoid potential issues with positivity violations. This resulted in a range of 100-190 mmHg. To create the categorical variables from baseline and time updated SBP, 4 categories were generated based on evenly spaced percentiles, resulting in categories of 100-129 mmHg, 130-139 mmHg, 140-149 mmHg and >=150 mmHg.

8.2.4.3 Aspirin

Use of aspirin in the previous year, as defined by presence of a prescription in the patient's record (see appendix 12) was also added as an indicator for elevated CVD risk. This was entered both as a baseline and as a time updated variable.

8.2.4.4 Cancer

Occurrence of cancer during follow up may affect treatment decisions and may have an impact on risk of CVD and mortality, therefore it was considered important to include this as an adjustment. Patients with a history of cancer at diabetes diagnosis were excluded as described in 5.2.3. If a patient developed cancer during follow up, this was adjusted for with an indicator variable that was updated from 0 to 1 at the occurrence of their first cancer diagnosis. Consistent with the study entry requirements (as described in 4.4.2.1 and Figure 4.5), if a patient developed cancer between diabetes diagnosis and study entry they were not excluded. This is because if they had been followed up from diagnosis, this cancer occurrence would not be a censoring

event. This meant that only patients who were untreated at study entry could have a history of cancer at study entry. Since this raises possible issue of selection bias, a sensitivity analysis was also performed where these patients were excluded.

8.2.5 Analysis plan

8.2.5.1 Descriptive analyses

Demographics at study entry for the updated cohort were presented by medication at study entry, as previously described (see 7.2.7.1). Crude incidence rates of MI, stroke and all-cause mortality after study entry were calculated, both overall and per month of follow up. These were calculated after the data had been split into intervals, by fitting a logistic regression to approximate the incidence per 1000 person years. To gain an understanding of possible HbA1c trajectories, a sample of observed trajectories for different treatment patterns was plotted. To gain an understanding of the extent to which confounding by the selected covariates may actual be present in the data, associations between all outcomes and all covariates (in categorical form) were estimated, considering each covariate in turn. These were adjusted for age, gender, smoking status and a time updating indicator for diabetes treatment (none, metformin, sulfonylureas). This ensured that associations with outcome were not solely due to a mediating effect of diabetes treatment, or due to confounding by strong known risk factors. For HbA1c, BMI and SBP, the association between baseline and most recent values and outcome were estimated with both variables (again in their categorical form) in a single model since it was anticipated that change from baseline may be important to capture.

Modelling only the most recent value for a time-dependent confounder assumes that the rest of the covariate history has no independent association with the risk of outcome. It was felt that shorter term changes in HbA1c and SBP not captured by the change from baseline may also be predictive of CV outcomes, so to capture this, the values of these covariates two intervals back were also investigated for their association with outcome. Because it is likely that changes in HbA1c and SBP are also associated with each other, the effects of these additional variables were estimated from a single model, adjusting for age, gender, smoking, time updated diabetes medications, and the baseline and most recent values SBP and HbA1c. Due to concerns over collinearity from using a LOCF approach, these additional variables were only added to the weighting models if there was a strong suggestion of association with the outcome. Specifically, they were included if the variable was statistically significant ($p<0.05$), or the estimated HR for

any level of the covariate vs the reference level were >1.15 or <0.85. Short term change in BMI was not considered for the primary analysis as overall, as BMI changes much more slowly (as shown in 6.3.2) and therefore concerns over collinearity in the LOCF approach were greater. However, since rapid decline in BMI could be a strong indicator of increasing frailty, or could be a result of bariatric surgery, a time updated variable to capture short term change in BMI was added to the weighting model in a sensitivity analysis (see 8.2.5.3).

8.2.5.2 Primary analyses

8.2.5.2.1 Models for IPTW and IPCW

IPTW were estimated as described in 4.4.2. In contrast to the previous chapter, initiation of sulfonylureas when interested in the effect of metformin was not considered a censoring event, and vice versa. Therefore it was possible to estimate the probability of treatment with metformin or sulfonylureas within the same model, using multinomial logistic regression with three outcome levels.

As with the previous analysis, continuous variables were entered into the model with varying levels of complexity. Since in the previous chapter the simplified form of the spline model was not observed to improve the weighting model, it was decided to compare just two different covariate specifications. Specification A used natural cubic splines for all HbA1c, BMI and SBP variables, as well as age, time since study entry and time between diabetes diagnosis and study entry. Categorical/binary variables were used for all other covariates. Specification B was the same as A, but with categorical variables for age, and all HbA1c, BMI and SBP variables. A simplified continuous form for time since study entry (simplified from the full spline as previously described in 7.2.7.2.2) was also used.

Although it was not expected for the overall shape of associations between continuous covariates and probability of treatment to be different from the previous chapter, the optimal parameterisations of the cubic splines were re-estimated using the MVRS function [202]. This was done because of the restriction to 10 years follow up, the addition of new continuous variables, and to have a parameterisations that could appropriately model both the association with metformin initiation and sulfonylurea initiation. The resulting estimated spline parameterisations are presented in appendix 20.

As in the previous chapter (see 7.2.6), if a patient was censored for death, transfer out or change in medication, they were censored at the beginning of the interval. Unlike the previous chapter, there is a greater risk that this could result in missing fatal stroke or MI events; or non fatal events that resulted in the need for intensification of treatment. The potential impact of this was assessed in a sensitivity analysis (see 8.2.5.3.4). IPCW were estimated as described in 4.4.1.3, using a multinomial logistic regression to model the probabilities for censoring due to death, transfer from practice, or initiation of any other medication other than metformin or a sulfonylurea. Initial investigations suggested that the new parameterisation of baseline BMI estimated by the MVRS function was not appropriate to model the association between baseline BMI and probability of death (see appendix 21). Therefore for covariate specification A, baseline BMI was entered into the censoring model as a natural cubic spline with knots at 20, 27, 30 and 50. This was the parameterisation used in the previous chapter (see Table 7.5), which was more flexible. If the joint IPTW and IPCW models made little difference to the main analyses compared to IPTW alone (as defined by a relative change in HR of <10%), the addition of IPCW would be dropped for the sensitivity analysis where appropriate. This was because the addition of IPCW doubled the time taken to estimate the weights, so removing them for multiple sensitivity analyses was more efficient.

8.2.5.2.2 Outcome models

A single model was fitted per outcome. The effects of metformin vs diet only and sulfonylureas vs diet only on risk of MI, stroke and all-cause mortality were estimated using a single pooled logistic regression model with robust standard errors. Longitudinal HbA1c was modelled using a GEE with independent working correlation matrix. Both of these methods have been described previously (see section 4.4). For longitudinal HbA1c, because time since study entry was entered in both model specifications as a cubic spline, no assumption was made about the shape of the trajectory of HbA1c. For this model, as well as including all baseline covariates, an interaction term between baseline HbA1c and time since study entry was included. This was done to allow for the expected change in HbA1c through time to depend upon the starting value independently from any treatment effect [222].

To investigate the effects of the two treatments though time, three alternative parameterisations of exposure were used. Firstly, a three level exposure for current treatment was used (none, metformin, sulfonylureas). This assumed a constant effect of treatment, and

was used to allow a direct comparison to the results of the UKPDS study and other existing research that estimate a single hazard ratio (HR) from a cox PH model. For longitudinal HBA1c, this assumes that absolute differences in HbA1c remain fixed through time, so the model assumes three parallel (but possibly non-linear) trajectories for the three treatment options. Secondly, since current treatment may not completely capture full treatment history, time since first prescription (assumed to represent cumulative medication use) was calculated. This was entered into the outcome model as a categorical variable, with groupings as follows: none, 1-3 months metformin, 3-6 months metformin, 6-12 months metformin, 1-2 years metformin, 2-5 years metformin, >5 years metformin, 1-3 months sulfonylureas, 3-6 months sulfonylureas, 6-12 months sulfonylureas and so on. Broader categorisations later in time were to reduce sparsity and because differences of a few months would be less important after longer exposure. To increase flexibility, cumulative medication was also modelled using a natural cubic spline with five knots at 0, 6, 12, 40 and 100 months, and entered into the model as interaction with the 3 level variable for current treatment to allow separate curves to be estimated for metformin and sulfonylureas. This model also included the three level variable for current treatment as a separate variable to allow an immediate change in treatment between 0 months of exposure and 1 month of exposure, which may not be fully captured by a smooth continuous function. For the time-to-event outcomes, modelling time on treatment in this way allows for non-proportional hazards. For longitudinal HbA1c, this approach allows for the difference in HbA1c trajectories between the treatments to vary through time.

For each exposure definition, five models were fitted to estimate the effect of metformin and sulfonylureas. As in the previous chapter, this was to investigate the impact of the different modelling approaches. Three models were the “standard analyses” (1: basic baseline adjustment, 2: full baseline adjustment, 3: full baseline adjustment plus time updated covariates) and two were the weighted MSMs (4: IPTW only and 5: joint IPTW/IPCW). For the repeated measures outcome, the third standard analysis model was not fitted because time updated values of HbA1c are the outcome, and such a model would not have a clear interpretation.

Consistent with the previous chapter, the parametrisations of the baseline covariates for each model specification were the same as those entered into the treatment model. This was done to simplify code and to minimise the size of the data as far as possible during the computation process. As previously discussed (see 7.4.3.3), the same spline parameterisation does not enforce the same shape of association as the treatment model, but it if the parametrisation is too simple for the association between the covariate and outcome, there is the risk of

misspecifying the outcome model. To investigate whether this may be the case, initial checks were performed to compare the estimated associations between baseline covariates and each outcome using the parameterisations used in the treatment model, and parameterisations of 4 knots based at the 20th, 40th, 60th and 80th percentiles of the variable's distribution. This was done in an unweighted model with full baseline adjustment (model 2). All comparisons are shown graphically in appendix 21. With the exception of the association between baseline BMI and all-cause mortality, no appreciable differences were observed to suggest that using the parameterisations from the treatment model would cause serious model misspecification. Baseline BMI was entered into the outcome model for all–cause mortality in a more flexible form because of differences observed in this exploratory work. Specifically, it was entered as a natural cubic spline with knots at 20, 27, 30 and 50 kg/m^2, which was observed to better model the apparent non-linear association, as shown in appendix 21. For clarity, the exact spline parameterisations and categorisation used are tabulated in appendix 20.

8.2.5.3 Sensitivity analyses

All sensitivity analyses were performed using model specification A only (cubic splines for time, HbA1c, BMI, SBP and age), and with categorical cumulative exposure unless otherwise stated. Since the main aim was to establish if the estimates from the MSM were affected, the sensitivity analyses were restricted to the weighted models only. As explained in 8.2.5.2.1, IPCW to adjust for loss to follow up by death, transfer out or intensification to a medication other than metformin or sulfonylureas, were not included in the sensitivity analyses unless specifically stated.

8.2.5.3.1 Restrict population to one more similar to UKPDS population.

The UKPDS study was conducted among "overweight" patients (defined being >120% of ideal body weight), and recruited patients aged 25-65 years. This sensitivity analysis aimed to recreate this population as far as possible, by restricting to patients aged < 65 years at time of diabetes diagnosis, (though the original lower bound for the cohort remained 32 years) and that had a BMI of 25 kg/m^2 or more at study entry.

8.2.5.3.2 Exclude any history of cancer before study entry

The analysis was repeated excluding patients that had a history of cancer between time of diagnosis and study entry, to check whether there was any suggestion of selection bias from their inclusion.

8.2.5.3.3 For MI and stroke, restrict analysis to those with no history of outcome.

This sensitivity analysis was done to assess the impact of allowing patients already at higher risk of the outcome to be included in the primary analysis.

8.2.5.3.4 For MI and stroke, do not censor for death or treatment intensification at the beginning of the interval if an event occurs in the same interval.

This analysis altered the censoring such that the occurrence of an MI or stroke event and death in the same interval was re-coded as an outcome event instead of a censoring event. The sensitivity analysis was then extended to additionally not censor if a patient experienced their MI or Stroke in the same interval in which they intensified treatment.

8.2.5.3.5 Censor at any change from first initiated treatment.

This sensitivity analysis was designed to assess whether assuming an intention to treat principle until intensification with something other than sulfonylureas or metformin resulted in a different estimate of treatment effect compared to an as treated analysis. IPCW were used for this analysis, with the addition of sulfonylurea initiation as a censoring event for the metformin model, and vice versa. Therefore, the IPTW were re-estimated with separate models for metformin and sulfonylureas as in chapter 7. Since this analysis approach was most likely to affect the treatment estimates for longer follow up, the weighted outcome model fitted was that with cumulative medication in its spline form.

8.2.5.3.6 Remove TIA from definition of stroke

The definition of stroke in the main analysis included TIA. To look at the potential impact of this, TIA was excluded from the definition of stroke and prior TIA was included as a time-dependent predictor in the weighting model.

8.2.5.3.7 Addition of short term change in BMI to weighting models

The primary analysis did not capture changes in BMI other than that from baseline to the current value, since BMI typically changes slowly. However, it is possible that rapid weight decline may be predictive of mortality. Conversely, it may also indicate bariatric surgery which may reduce risk of CV events. To check whether the omission of this variable had an impact on the estimated treatment effects for the time to event outcomes, the IPTW's were re-estimated including BMI two intervals back as an additional time-dependent variable. Since this variable was not included originally, no spline parameterisation was estimated in advance. For simplicity, a natural cubic spline with 4 evenly spaced knots at the 20^{th}, 40^{th}, 60^{th} and 80^{th} percentiles (equating to 22, 28, 33 and 42 kg/m^2) was used.

8.3 RESULTS

8.3.1 Cohort description

The updated cohort of patients with incident type 2 diabetes consisted of 57,675 subjects alive and eligible to be included in the analysis at baseline (study entry). Demographics are presented by medication at baseline in Table 8.1.

	No Medication			Metformin			Sulfonylureas		
	Mean	**SD**	**Median, IQR**	**Mean**	**SD**	**Median, IQR**	**Mean**	**SD**	**Median, IQR**
Age at diabetes diagnosis	62.3	12.0	63 (54 - 71)	57.6	11.8	57 (49 - 66)	59.5	13.3	59 (49 - 70)
Time between diagnosis and study entry (months)	3.4	9.8	0 (0 - 3)	0.0	0.0	0 (0 - 0)	0.0	0.0	0 (0 - 0)
A1c at study entry	7.3	1.6	6.8 (6.3 - 7.7)	9.3	2.2	8.9 (7.4 - 11)	10.9	2.2	11.1 (9.1 - 12.7)
BMI at study entry	31.7	6.1	30.8 (27.4 - 35.1)	33.4	6.5	32.4 (28.7 - 37.2)	27.6	5.5	26.2 (23.8 - 30.2)
SBP at study entry	137.8	16.3	138 (128 - 147)	136.9	16.0	136 (127 - 146)	136.5	17.6	135 (124 - 147)
	N	**%**		**N**	**%**		**N**	**%**	
Sex									
Male	27,799	56		4,130	59		596	62	
Female	21,951	44		2,832	41		367	38	
History of cancer *									
No	49,569	99.6		6,962	100		963	100	
Yes	181	0.4		0	0		0	0	
History of CVD									
No	42,183	85		6,230	89		860	89	
Yes	7,567	15		732	11		103	11	
CVD event in past 3 months									
No	48,794	98		6,890	99		943	98	
Yes	956	2		72	1		20	2	
History of MI									
No	47,925	96		6,771	97		930	97	
Yes	1,825	4		191	3		33	3	
MI in past 3 months									
No	49,588	99.7		6,950	99.8		956	99	
Yes	162	0.3		12	0.2		7	1	
History of Stroke									
No	47,949	96		6,804	98		933	97	
Yes	1,801	4		158	2		30	3	
Stroke in past 3 months									
No	49,550	99.6		6,944	99.7		959	99.6	
Yes	200	0.4		18	0.3		4	0.4	
History of CKD									
No	46,436	93		6,673	96		888	92	
Yes	3,314	7		289	4		75	8	

Table 8.1 Cohort demographic at time of study entry,stratified by medication at study entry

*only patients with delayed study entry are able to have a history of cancer at study entry therefore this is only patients who are not on treatment at study entry (as already explained in 8.2.4.4)

	No Medication		Metformin		Sulfonylureas	
	N	**%**	**N**	**%**	**N**	**%**
Use of statins in previous year						
No	23,598	47	3,124	45	571	59
Yes	26,152	53	3,838	55	392	41
Use of anti HTs in previous year						
No	17,765	36	3,174	46	523	54
Yes	31,985	64	3,788	54	440	46
Use of NSAIDS in previous year						
No	40,056	81	5,783	83	802	83
Yes	9,694	19	1,179	17	161	17
Use of ASPIRIN in previous year						
No	35,285	71	5,445	78	727	75
Yes	14,465	29	1,517	22	236	25
Smoking Status						
non	19,942	40	2,754	40	396	41
current	8,764	18	1,474	21	238	25
ex	21,044	42	2,734	39	329	34
Alcohol consumption						
non-drinker	5,667	11	952	14	105	11
ex-drinker	3,797	8	660	9	75	8
current drinker unknown	882	2	129	2	25	3
rare drinker <2u/d	11,951	24	1,761	25	222	23
moderate drinker 3-6u/d	22,648	46	2,904	42	443	46
excessive drinker >6u/d	4,805	10	556	8	93	10
Year of diabetes onset						
2000-2005	12,315	25	578	8	284	29
post 2005	37,435	75	6,384	92	679	71

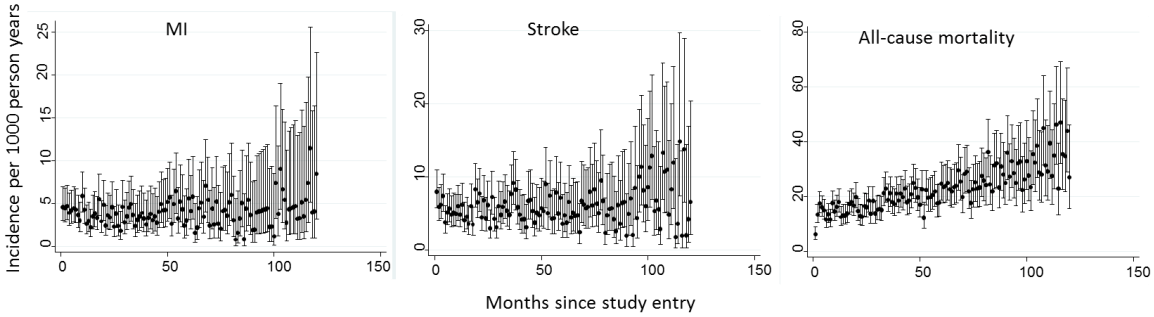
Table 8.1 continued: Cohort demographics of at time of study entry continued,stratified by medication at study entry

8.3.2 Descriptive analysis of outcomes

Total follow up time for the 57,675 patients was 259,660 person years for MI, 257,781 person years for stroke, and 263,036 person years for all-cause mortality. There were 1007 MI events, 1471 stroke events and 5121 deaths observed during follow up. Crude incidences (per 1000 person years) with 95% CI were 3.9 (3.6 – 4.1), 5.7 (5.4-6.0) and 19.5 (19.0-20.0) for MI, stroke and all-cause mortality respectively.

Crude Incidence per 1000 person years of MI, stroke and all-cause mortality for each month of follow up and are presented in Figure 8.1.

Figure 8.1 Crude incidence (per 1000 person years) of MI, Stroke and all-cause mortality per month of follow up*

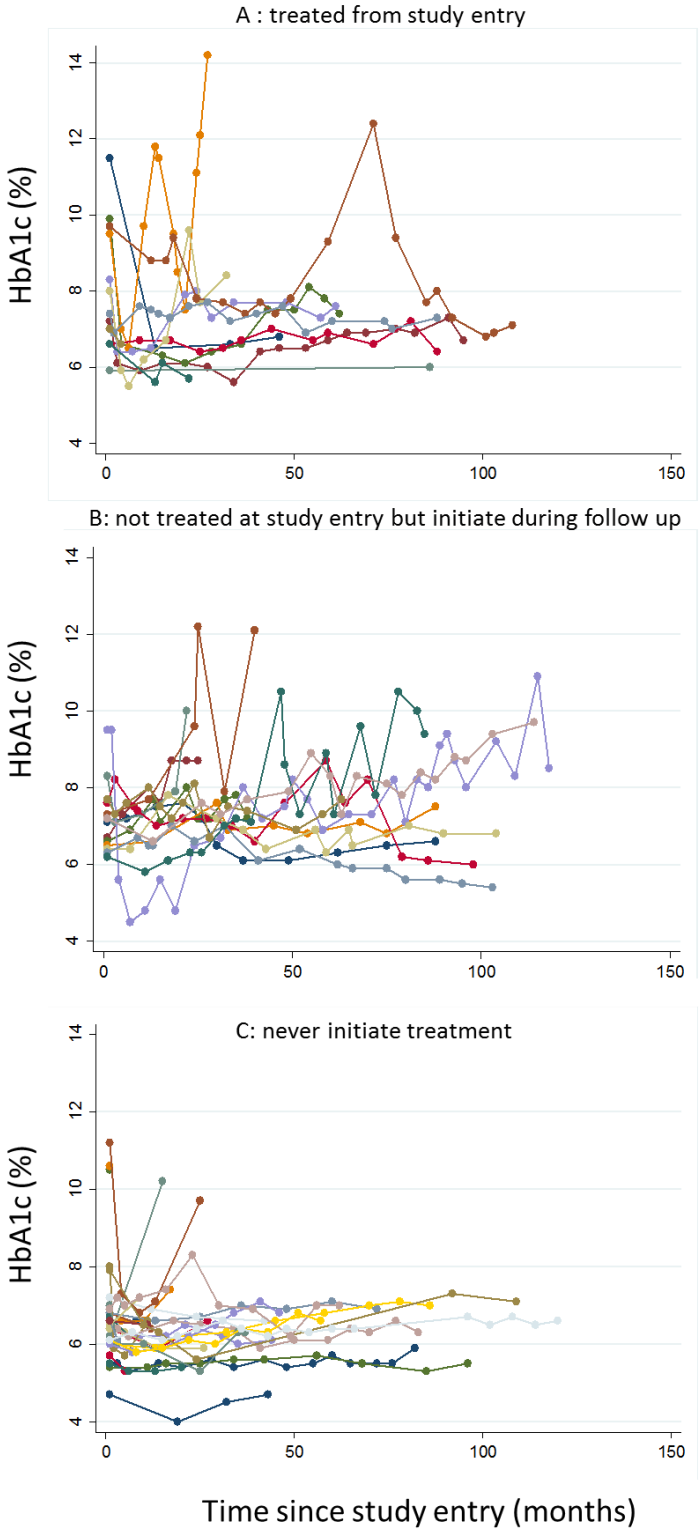


*estimated using pooled logistic regression with month since study entry as a categorical variable with 120 levels (one per month). Bars represent 95% confidence intervals.

The frequency of HbA1c measures was highly variable between patients; as previously described in chapter 6. Figure 8.2 displays HbA1c through time for a selection of individuals who had different medication patterns from study entry. There appeared to be no clear pattern in trajectory based on medication decisions alone. HbA1c tended to fluctuate frequently, particularly in those who initiated treatment at some point (Figure 8.2A and B).

Associations between covariates and the three time to event outcomes are displayed in Table 8.2. The estimated associations with longitudinal HbA1c are given in Table 8.3. The majority of covariates that were identified in advance to be included in the weighting models showed at least some association with all outcomes. The additional term for HbA1c two intervals back met the criteria for inclusion (see 8.2.5.1) for all outcomes and so was added to the models for treatment and censoring weights. However, the additional value for SBP two interval back did not appear to add further information for any outcome and so was not included. The results suggested that same weighting model (in terms of covariate selection) could be assumed appropriate for all outcomes.

Figure 8.2 Sample of observed longitudinal HbA1c in patients who A were treated from study entry, B Introduced treatment part way through and C never started treatment.



Different colours represent different individual patients

.

	MI		Stroke		All-cause mortality	
Risk Factor	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**
Age						
<45	1(ref)		1(ref)		1(ref)	
45-59	2.23	(1.50 , 3.31)	2.79	(1.82 , 4.28)	2.37	(1.80 , 3.13)
60-74	3.24	(2.20 , 4.76)	5.63	(3.71 , 8.54)	7.23	(5.53 , 9.45)
75+	6.62	(4.47 , 9.81)	12.33	(8.09 , 18.77)	23.60	(18.03 , 30.90)
Gender						
Male	1(ref)		1(ref)		1(ref)	
Female	0.54	(0.47 , 0.61)	0.9	(0.81 , 1.00)	0.84	(0.79 , 0.89)
Smoking status						
Non	1(ref)		1(ref)		1(ref)	
Current	2.19	(1.85 , 2.59)	1.46	(1.26 , 1.69)	2.51	(2.32 , 2.71)
Ex	1.38	(1.20 , 1.60)	1.18	(1.05 , 1.32)	1.55	(1.45 , 1.65)
Drinking Status						
Non drinker	1(ref)		1(ref)		1(ref)	
Ex-drinker	1.60	(1.19 , 2.14)	1.04	(0.82 , 1.32)	1.25	(1.11 , 1.42)
Current drinker unknown	1.88	(1.23 , 2.90)	1.03	(0.69 , 1.53)	1.55	(1.29 , 1.86)
Rare drinker <2u/d	1.35	(1.07 , 1.71)	0.95	(0.79 , 1.13)	0.94	(0.86 , 1.04)
Moderate drinker 3-6u/d	1.21	(0.97 , 1.51)	0.92	(0.78 , 1.09)	0.89	(0.81 , 0.97)
Excessive drinker >6u/d	1.09	(0.81 , 1.48)	0.91	(0.72 , 1.16)	1.22	(1.08 , 1.38)
Year of diabetes onset						
2000-2005	1(ref)		1(ref)		1(ref)	
Post 2005	0.98	(0.85 , 1.12)	0.77	(0.69 , 0.86)	0.95	(0.90 , 1.01)
Use in year before baseline…						
Anti-hypertensives	1.55	(1.34 , 1.81)	1.32	(1.17 , 1.50)	1.18	(1.10 , 1.26)
Statins	1.36	(1.19 , 1.55)	1.13	(1.02 , 1.26)	0.95	(0.90 , 1.01)
NSAIDS	1.20	(1.04 , 1.39)	1.13	(1.00 , 1.28)	0.92	(0.86 , 0.99)
Aspirin	1.73	(1.52 , 1.97)	1.40	(1.26 , 1.56)	1.19	(1.12 , 1.26)
Event in three months before baseline of…						
Any CVD	2.82	(2.19 , 3.63)	1.76	(1.36 , 2.27)	1.27	(1.09 , 1.47)
MI	5.79	(3.52 , 9.52)	0.62	(0.20 , 1.94)	1.77	(1.22 , 2.57)
Stroke	2.46	(1.35 , 4.47)	5.63	(3.97 , 7.99)	1.46	(1.06 , 2.01)
Any history at baseline of …						
CVD	2.59	(2.26 , 2.97)	1.94	(1.73 , 2.18)	1.51	(1.42 , 1.61)
MI	3.56	(2.91 , 4.35)	1.33	(1.05 , 1.68)	1.69	(1.52 , 1.90)
Stroke	1.56	(1.20 , 2.01)	3.51	(2.99 , 4.12)	1.42	(1.27 , 1.58)
CKD	1.21	(0.95 , 1.54)	1.10	(0.90 , 1.34)	1.45	(1.32 , 1.59)
Cancer	1.06	(0.38 , 3.00)	0.60	(0.22 , 1.68)	2.27	(1.67 , 3.09)
Time updated use in previous year of..						
Anti-hypertensives	1.85	(1.55 , 2.22)	1.58	(1.36 , 1.83)	0.90	(0.84 , 0.97)
Statins	1.06	(0.92 , 1.23)	1.15	(1.02 , 1.3)	0.71	(0.67 , 0.75)
NSAIDS	1.04	(0.89 , 1.23)	1.08	(0.94 , 1.23)	1.08	(1.00 , 1.16)
Aspirin	1.82	(1.60 , 2.07)	1.79	(1.61 , 1.99)	1.14	(1.08 , 1.21)
Event in previous three months of…						
Any CVD	6.82	(5.01 , 9.27)	3.15	(2.15 , 4.62)	3.16	(2.66 , 3.76)
MI	71.78	(30.92 , 166.63)	2.03	(0.76 , 5.44)	4.32	(3.08 , 6.06)
Stroke	2.27	(0.94 , 5.47)	27.12	(10.05 , 73.21)	3.94	(2.99 , 5.18)
Time updated history of …						
CVD	2.80	(2.45 , 3.20)	2.10	(1.88 , 2.35)	1.71	(1.61 , 1.81)
MI			1.36	(1.11 , 1.68)	1.88	(1.71 , 2.07)
Stroke	1.61	(1.29 , 2.00)			1.69	(1.55 , 1.85)
CKD	1.21	(1.03 , 1.43)	1.02	(0.89 , 1.17)	1.51	(1.41 , 1.61)
Cancer	1.14	(0.85 , 1.53)	1.03	(0.81 , 1.32)	4.70	(4.32 , 5.11)

Table 8.2 Estimated HR and 95% CI for the association between covariates and outcomes of MI, stroke and all-cause mortality.

Each covariate considered separately, with adjustment for age, smoking status and time updated diabetes medication. HR approximated from a pooled logistic regression with time since baseline as underlying time scale.

	MI		Stroke		All-cause mortality	
Risk Factor	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**
[a]Baseline BMI						
<25	1(ref)		1(ref)		1(ref)	
25-29	1.02	(0.79 , 1.32)	0.75	(0.61 , 0.92)	1.22	(1.11 , 1.34)
30-34	0.90	(0.66 , 1.23)	0.82	(0.63 , 1.06)	2	(1.76 , 2.27)
35+	0.74	(0.49 , 1.11)	0.77	(0.55 , 1.08)	3.06	(2.56 , 3.65)
[a]Baseline HbA1c						
<6%	1(ref)		1(ref)		1(ref)	
6-6.5%	1.04	(0.81 , 1.33)	0.95	(0.78 , 1.14)	1.14	(1.03 , 1.26)
6.5 -7%	1.25	(0.98 , 1.61)	0.87	(0.72 , 1.06)	1.27	(1.14 , 1.41)
7-8%	1.11	(0.86 , 1.44)	0.89	(0.72 , 1.08)	1.4	(1.25 , 1.57)
8-10%	1.10	(0.83 , 1.46)	0.84	(0.67 , 1.05)	1.42	(1.25 , 1.61)
10% +	1.21	(0.89 , 1.64)	0.59	(0.45 , 0.77)	1.21	(1.05 , 1.40)
[a]Baseline SBP						
100-129	1(ref)		1(ref)		1(ref)	
130-139	1.07	(0.89 , 1.28)	0.95	(0.82 , 1.11)	0.91	(0.84 , 0.98)
140-149	1.00	(0.83 , 1.20)	1.03	(0.89 , 1.20)	0.96	(0.89 , 1.04)
150+	1.02	(0.84 , 1.23)	1.19	(1.03 , 1.37)	1.01	(0.93 , 1.10)
[b]Previous BMI (t -1)						
<25	1(ref)		1(ref)		1(ref)	
25-29	1.06	(0.84 , 1.34)	1.12	(0.92 , 1.35)	0.39	(0.36 , 0.43)
30-34	1.17	(0.87 , 1.57)	1.06	(0.83 , 1.37)	0.23	(0.21 , 0.27)
35+	1.12	(0.76 , 1.66)	0.91	(0.65 , 1.29)	0.18	(0.15 , 0.21)
[b]Previous HbA1c (t -1)						
<6%	1(ref)		1(ref)		1(ref)	
6-6.5%	0.65	(0.35 , 1.22)	0.81	(0.46 , 1.40)	0.62	(0.50 , 0.78)
6.5 -7%	0.33	(0.18 , 0.6)	0.85	(0.43 , 1.68)	0.55	(0.42 , 0.73)
7-8%	0.39	(0.21 , 0.72)	1.11	(0.56 , 2.18)	0.54	(0.40 , 0.74)
8-10%	0.38	(0.17 , 0.83)	0.99	(0.47 , 2.11)	0.54	(0.37 , 0.78)
10% +	0.38	(0.15 , 1.00)	1.36	(0.53 , 3.49)	0.56	(0.33 , 0.97)
[b]Previous SBP (t -1)						
100-129	1(ref)		1(ref)		1(ref)	
130-139	0.95	(0.67 , 1.35)	0.92	(0.8 , 1.06)	0.6	(0.52 , 0.68)
140-149	1.06	(0.73 , 1.55)	1.04	(0.91 , 1.20)	0.53	(0.45 , 0.62)
150+	1.28	(0.85 , 1.93)	1.51	(1.3 , 1.75)	0.58	(0.49 , 0.69)
[c]Previous HbA1c (t -2)						
<6%	1(ref)		1(ref)		1(ref)	
6-6.5%	1.56	(0.81 , 2.97)	1.16	(0.67 , 2.02)	1.03	(0.82 , 1.29)
6.5 -7%	2.77	(1.50 , 5.10)	1.08	(0.55 , 2.11)	1.02	(0.77 , 1.35)
7-8%	2.88	(1.50 , 5.53)	0.97	(0.49 , 1.92)	1.10	(0.80 , 1.51)
8-10%	3.52	(1.59 , 7.79)	1.25	(0.59 , 2.67)	1.33	(0.92 , 1.93)
10% +	4.95	(1.88 , 13.05)	1.44	(0.56 , 3.70)	1.48	(0.86 , 2.54)
[c]Previous SBP (t-2)						
100-129	1(ref)		1(ref)		1(ref)	
130-139	0.91	(0.65 , 1.29)	1.12	(0.84 , 1.49)	0.91	(0.79 , 1.04)
140-149	0.90	(0.62 , 1.31)	0.85	(0.62 , 1.16)	0.91	(0.78 , 1.06)
150+	0.96	(0.63 , 1.45)	0.93	(0.66 , 1.31)	1.02	(0.87 , 1.21)

Table 8.2 cont. Estimated HR and 95% CI for the association between covariates and outcomes of MI, Stroke and all-cause mortality.

Each covariate considered separately, with adjustment for age, smoking status and time updated diabetes medication. HR approximated from a pooled logistic regression with time since baseline as underlying time scale.
[a] Additionally adjusted for value of covariate in previous interval. [b] Additionally adjusted for value of covariate at baseline [c] Estimated from a single model containing baseline HbA1c, previous HbA1c (interval t-1) , baseline SBP, and previous SBP (interval t-1).

Risk Factor	Difference in HbA1c	95% CI
Age (years)		
<45	0(ref)	
45-59	-0.21	(-0.25 , -0.17)
60-74	-0.42	(-0.45 , -0.38)
75+	-0.47	(-0.51 , -0.44)
Gender		
Male		
Female	-0.02	(-0.03 , 0.00)
Smoking status		
Non	0(ref)	
Current	0.08	(0.06 , 0.11)
Ex	0.00	(-0.01 , 0.02)
Drinking Status		
Non drinker	0(ref)	
ex-drinker	-0.04	(-0.07 , 0.00)
current drinker unknown	0.11	(0.04 , 0.18)
rare drinker <2u/d	0.00	(-0.03 , 0.03)
moderate drinker 3-6u/d	-0.03	(-0.05 , 0.00)
excessive drinker >6u/d	-0.13	(-0.16 , -0.10)
Year of diabetes onset		
2000-2005		
post 2005	-0.02	(-0.04 , 0.00)
Use in year before baseline of…		
Anti-hypertensives	-0.05	(-0.07 , -0.04)
Statins	-0.02	(-0.03 , 0.00)
NSAIDS	0.01	(-0.01 , 0.03)
Aspirin	0.01	(-0.01 , 0.02)
Event in three months before baseline of…		
Any CVD	0.00	(-0.05 , 0.05)
MI	0.00	(-0.12 , 0.13)
Stroke	-0.04	(-0.14 , 0.06)
Any history at baseline of …		
CVD	0.06	(0.04 , 0.08)
MI	0.08	(0.04 , 0.12)
Stroke	0.01	(-0.03 , 0.04)
CKD	-0.05	(-0.08 , -0.02)
Cancer	-0.03	(-0.16 , 0.10)
Time updated use in previous year of..		
Anti-hypertensives	-0.07	(-0.09 , -0.06)
Statins	-0.05	(-0.07 , -0.04)
NSAIDS	-0.03	(-0.04 , -0.01)
Aspirin	-0.03	(-0.05 , -0.02)
Event in previous three months of…		
Any CVD	0.02	(-0.02 , 0.06)
MI	0.05	(-0.06 , 0.16)
Stroke	0.01	(-0.06 , 0.09)
Time updated history of …		
CVD	0.05	(0.03 , 0.07)
MI	0.06	(0.03 , 0.10)
Stroke	0.00	(-0.03 , 0.03)
CKD	-0.08	(-0.1 , -0.06)
Cancer	-0.01	(-0.05 , 0.03)

Risk Factor	Difference in HbA1c	95% CI
[a]Baseline BMI (kg/m²)		
<25	0(ref)	
25-29	0.04	(0.02 , 0.07)
30-34	0.09	(0.07 , 0.12)
35+	0.12	(0.09 , 0.15)
[a]Baseline HbA1c		
<6%	0(ref)	
6-6.5%	0.10	(0.09 , 0.11)
6.5 -7%	0.13	(0.12 , 0.14)
7-8%	0.17	(0.16 , 0.18)
8-10%	0.24	(0.22 , 0.25)
10% +	0.33	(0.31 , 0.34)
[a]Baseline SBP (mmHg)		
100-129	0(ref)	
130-139	-0.02	(-0.03 , -0.01)
140-149	-0.03	(-0.04 , -0.03)
150+	-0.06	(-0.07 , -0.05)
[b]Previous BMI (t -1)		
<25	0(ref)	
25-29	0.10	(0.08 , 0.12)
30-34	0.21	(0.19 , 0.23)
35+	0.28	(0.26 , 0.31)
[b]Previous HbA1c (t -1)		
<6%	0(ref)	
6-6.5%	0.49	(0.49 , 0.50)
6.5 -7%	0.94	(0.93 , 0.95)
7-8%	1.49	(1.48 , 1.50)
8-10%	4.03	(4.00 , 4.06)
10% +		
[b]Previous SBP (t -1)		
100-129	0(ref)	
130-139	0.05	(0.03 , 0.07)
140-149	0.07	(0.05 , 0.09)
150+	0.12	(0.10 , 0.14)
[c]Previous HbA1c (t -2)		
<6%	0(ref)	
6-6.5%	0.05	(0.00 , 0.09)
6.5 -7%	0.09	(0.03 , 0.15)
7-8%	0.11	(0.04 , 0.17)
8-10%	-0.04	(-0.12 , 0.03)
10% +	-0.34	(-0.45, -0.23)
[c]Previous SBP (t-2)		
100-129	0(ref)	
130-139	-0.01	(-0.03 , 0.00)
140-149	-0.01	(-0.03 , 0.01)
150+	0.01	(-0.01 , 0.03)

[a] Additionally adjusted for value of covariate in previous interval. [b] Additionally adjusted for value of covariate at baseline.
[c] Estimated from a single model containing baseline HbA1c, previous HbA1c (interval t-1) , baseline SBP, and previous SBP (interval t-1).

Table 8.3 Estimated associations between covariates and longitudinal HbA1c.

Each covariate considered separately, with adjustment for age, smoking status and time updated diabetes medication. . Estimate and 95% CI relate to absolute expected difference in HbA1c. Estimated from a generalised estimating equation including only intervals in which HbA1c is observed.

8.3.3 Models for the IPTW and IPCW

Exact numbers contributing and follow up times were slightly different for the different outcomes, but the full multivariable models estimating probability of treatment with metformin or sulfonylureas and probability of censoring were similar for all outcomes. The full model output for both treatment and censoring models for follow up to all-cause mortality, using the categorical covariate specification, are presented in appendix 22 as an example. The distribution of the calculated stabilised and truncated weights (IPTW only and joint IPTW and IPCW) for each outcome model are presented in Table 8.4. In the categorical models, the means of the stabilised weights were lower, suggesting that the categorisations reduced, to an extent, issues with positivity violations. Having said this, the 99th percentiles of the distribution were similar, the maximum weights for the categorical specification were still large, and truncation was still necessary. Overall, the means of the stabilised weights were much smaller than were observed in the previous chapter. For example, a mean of 4.18 was observed for the stabilised IPTW for MI vs a mean of 11,871 in the previous analysis (see 7.3.2.2). Since the truncated weights had means much closer to one, these were the weights taken forward to use in the outcome models.

8.3.4 Outcome models

For MI, stroke and all-cause mortality, the estimates obtained by the two covariate specifications were very similar. For clarity of presentation, only the results of covariate specification A (continuous parameterisations of covariates) are presented here, with results for covariate specification B (categorical parameterisations of covariates) displayed in appendix 23.

MI

	Categorical				Spline			
	Stabilised IPTW	Stabilised IPTW Truncated	Joint IPTW IPCW	Joint IPTW IPCW Truncated	Stabilised IPTW	Stabilised IPTW Truncated	Joint IPTW IPCW	Joint IPTW IPCW Truncated
Mean	1.27	1.00	1.26	0.99	4.18	0.99	4.65	0.99
SD	15.66	0.95	14.78	0.95	635.93	0.90	736.10	0.92
1st	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
25th	0.66	0.66	0.65	0.65	0.68	0.68	0.67	0.67
50th	0.90	0.90	0.88	0.88	0.91	0.91	0.90	0.90
75th	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99th	7.16	7.16	7.12	7.12	6.93	6.93	7.11	7.11

Stroke

	Categorical				Spline			
	Stabilised IPTW	Stabilised IPTW Truncated	Joint IPTW IPCW	Joint IPTW IPCW Truncated	Stabilised IPTW	Stabilised IPTW Truncated	Joint IPTW IPCW	Joint IPTW IPCW Truncated
Mean	1.26	1.00	1.26	0.99	4.32	0.99	4.83	0.99
SD	15.40	0.95	14.55	0.94	671.93	0.90	782.31	0.91
1st	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
25th	0.66	0.66	0.65	0.65	0.68	0.68	0.67	0.67
50th	0.90	0.90	0.88	0.88	0.92	0.92	0.90	0.90
75th	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99th	7.16	7.16	7.09	7.09	6.94	6.94	7.04	7.04

All-cause mortality

	Categorical				Spline			
	Stabilised IPTW	Stabilised IPTW Truncated	Joint IPTW IPCW	Joint IPTW IPCW Truncated	Stabilised IPTW	Stabilised IPTW Truncated	Joint IPTW IPCW	Joint IPTW IPCW Truncated
Mean	1.27	1.00	1.26	0.99	4.28	0.99	4.75	0.98
SD	15.78	0.95	15.04	0.94	660.54	0.90	771.96	0.91
1st	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
25th	0.65	0.65	0.65	0.65	0.68	0.68	0.67	0.67
50th	0.90	0.90	0.88	0.88	0.91	0.91	0.90	0.90
75th	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99th	7.16	7.16	7.08	7.08	7.04	7.04	7.06	7.06

HbA1c

	Categorical				Spline			
	Stabilised IPTW	Stabilised IPTW Truncated	Joint IPTW IPCW	Joint IPTW IPCW Truncated	Stabilised IPTW	Stabilised IPTW Truncated	Joint IPTW IPCW	Joint IPTW IPCW Truncated
Mean	1.27	1.00	1.26	0.99	4.29	0.99	4.77	0.99
SD	15.71	0.95	14.85	0.95	661.04	0.90	768.05	0.92
1st	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
25th	0.65	0.65	0.65	0.65	0.68	0.68	0.66	0.66
50th	0.90	0.90	0.88	0.88	0.91	0.91	0.90	0.90
75th	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99th	7.16	7.16	7.13	7.13	7.00	7.00	7.13	7.13

Table 8.4 Distribution of stabilised IPTW and joint IPTW and IPCW

Separate treatment models fitted for each of the 4 outcomes due to differing overall follow up. Weights estimated from two models differing in specification by the form of continuous variables, which were modelled as cubic splines (A, right) or categorised (b, left).

8.3.4.1 MI

The MSM with IPTW and IPCW estimated an overall summary hazard ratio for metformin vs diet for risk of MI as 0.93, with 95% CI from 0.73–1.18. The MSM without the IPCW was almost identical (Table 8.5). For metformin, this estimate was slightly lower than the unweighted model with full baseline adjustment (HR 0.98 (0.83-1.15)) which was itself lower than the standard model with basic baseline adjustment for age, time since diagnosis, smoking, alcohol and year of diabetes onset (HR 1.11, 0.96-1.28), though all confidence intervals had similar ranges. When examining the effect of cumulative metformin use on risk of MI, both MSMs gave some suggestion of a small increased risk in the first 3 months. For greater than 5 years, that the risk of MI compared to diet only was reduced by 34% (HR 0.66, CI 0.42-1.03) (Table 8.6). The unweighted model with full baseline adjustment estimated a similar trend, but the increased risk in the first 3 months was greater and the decreased risk with >5 years use was not as large, suggesting that the weighting may be working as expected. Modelling cumulative metformin use as a continuous spline more clearly showed the differences between standard analysis and weighted modes (Figure 8.3). Particularly, the decreased risk of MI with long term metformin use estimated by the MSMs was much more apparent.

For sulfonylureas vs diet, the overall HR for risk of MI was estimated to be 1.00, with a 95% CI of 0.66-1.51 (MSM with IPTW and IPCW). There were no appreciable differences between the fully adjusted baseline model and the MSMs (Table 8.5). When looking at cumulative use of sulfonylureas, both standard analysis methods and MSMs estimated early use of a sulfonylurea to reduce the risk of MI, with the HR over time appearing to vary above and below of one, albeit with wide confidence intervals (Table 8.6). This variation around a HR of one was also apparent when modelling cumulative use as a continuous function (Figure 8.3). As with metformin, differences between MSMs with IPTW and joint IPTW and IPCW were minimal.

	Metformin			Sulfonylureas		
	HR	SE	95% CI	HR	SE	95% CI
1 Basic adjustment	1.10	0.08	0.96 , 1.27	1.26	0.16	0.98 , 1.62
2 Full baseline adjustment	0.98	0.08	0.83 , 1.15	1.02	0.14	0.77 , 1.34
3 Time updated adjustment	0.93	0.08	0.79 , 1.10	0.95	0.14	0.72 , 1.26
4 IPTW model	0.92	0.12	0.72 , 1.18	1.01	0.21	0.68 , 1.52
5 IPTW and IPCW model	0.93	0.11	0.73 , 1.18	1.00	0.21	0.66 , 1.51

Table 8.5 HR for risk of MI with current use of metformin (left) or sulfonylureas (right) compared to diet only.

Models 1-3 are standard pooled logistic regression with varying levels of adjustment for confounders. Model 4 is a MSM with IPTW, and model 5 a MSM with joint IPTW and IPCW.**Basic adjustment** – age, gender, calendar period of diabetes onset, smoking and alcohol **Full baseline adjustment** – Basic adjustment + baseline measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **Time updated adjustment** – Full baseline adjustment plus time-dependent measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **IPTW model** – full baseline adjustment plus stabilised IPTW, where denominator model for IPTW includes all variables in time updated adjustment, and all baseline variables in the numerator. **IPTW and IPCW – IPTW** model plus joint weights for censoring, where censoring models include all baseline and time updated variables in denominator and all baseline variables in numerator.

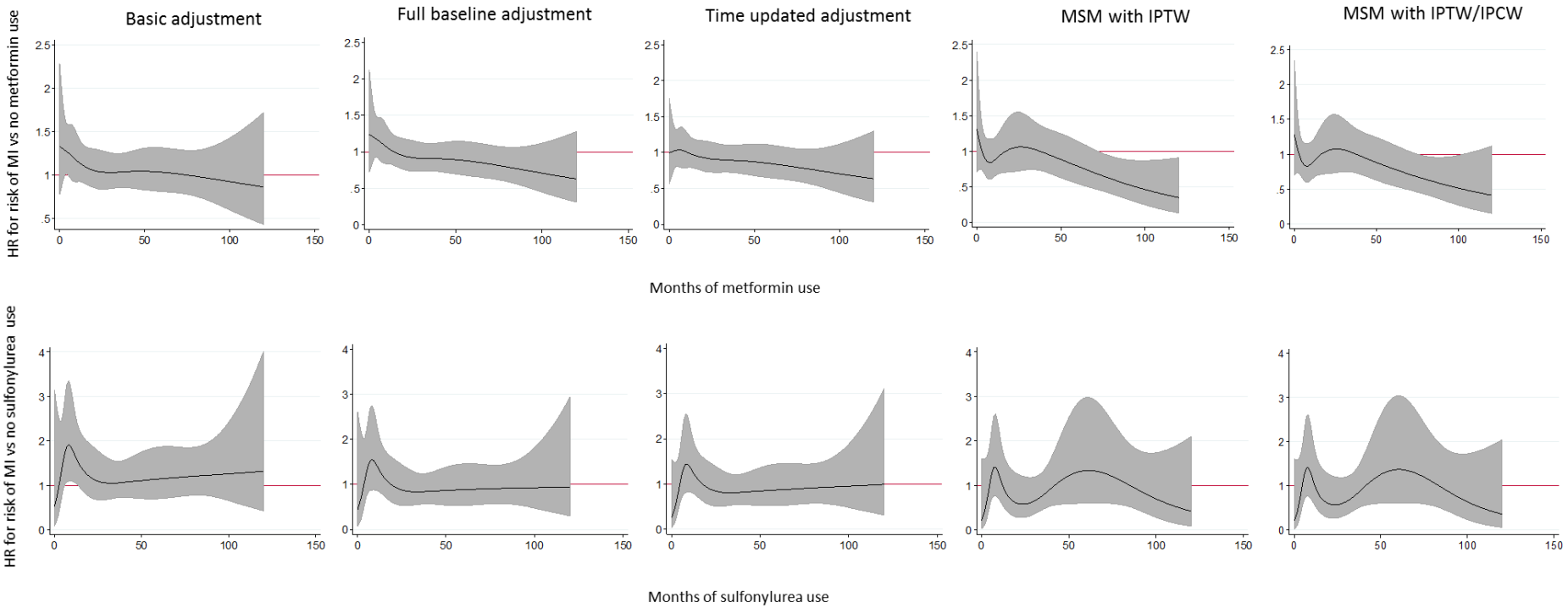
Metformin

	1 Basic Adjustment			2 Baseline Adjustment			3 Time updated adjustment			4 IPTW			5 IPTW and IPCW		
	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI
1- 3 months	1.30	0.23	0.91 , 1.84	1.20	0.22	0.84 , 1.71	1.00	0.19	0.69 , 1.46	1.13	0.24	0.74 , 1.72	1.10	0.24	0.72 , 1.68
3-6 months	1.19	0.23	0.82 , 1.73	1.09	0.21	0.75 , 1.61	0.97	0.19	0.65 , 1.43	0.90	0.21	0.57 , 1.41	0.88	0.21	0.56 , 1.39
6-12 months	1.18	0.18	0.88 , 1.58	1.07	0.16	0.8 , 1.45	1.02	0.16	0.75 , 1.38	0.81	0.15	0.57 , 1.15	0.81	0.14	0.57 , 1.15
1-2 years	1.16	0.14	0.92 , 1.47	1.04	0.13	0.81 , 1.34	1.01	0.13	0.79 , 1.30	1.10	0.23	0.73 , 1.67	1.11	0.24	0.73 , 1.69
2 - 5 years	1.05	0.11	0.86 , 1.28	0.91	0.10	0.73 , 1.13	0.89	0.10	0.72 , 1.11	0.94	0.16	0.68 , 1.30	0.94	0.15	0.68 , 1.30
> 5 years	0.88	0.13	0.67 , 1.17	0.71	0.11	0.52 , 0.97	0.69	0.11	0.51 , 0.94	0.62	0.14	0.4 , 0.98	0.65	0.15	0.42 , 1.01
	Sulfonylureas														
	1 Basic Adjustment			2 Baseline Adjustment			3 Time updated adjustment			4 IPTW			5 IPTW and IPCW		
	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI
1- 3 months	0.87	0.51	0.28 , 2.72	0.71	0.42	0.23 , 2.23	0.47	0.28	0.14 , 1.53	0.56	0.34	0.17 , 1.85	0.56	0.34	0.17 , 1.85
3-6 months	1.30	0.66	0.48 , 3.49	1.05	0.54	0.39 , 2.87	0.88	0.46	0.32 , 2.43	0.79	0.47	0.25 , 2.52	0.78	0.46	0.25 , 2.48
6-12 months	2.02	0.62	1.10 , 3.70	1.63	0.51	0.88 , 3.01	1.52	0.47	0.82 , 2.79	1.18	0.42	0.59 , 2.36	1.18	0.42	0.59 , 2.36
1-2 years	1.37	0.39	0.78 , 2.39	1.10	0.32	0.62 , 1.93	1.06	0.31	0.60 , 1.87	0.99	0.40	0.45 , 2.19	0.98	0.40	0.44 , 2.19
2 - 5 years	0.99	0.22	0.64 , 1.54	0.77	0.18	0.49 , 1.22	0.75	0.17	0.48 , 1.19	0.70	0.22	0.38 , 1.30	0.72	0.23	0.38 , 1.36
> 5 years	1.25	0.29	0.79 , 1.96	0.94	0.23	0.58 , 1.52	0.95	0.23	0.58 , 1.54	1.27	0.49	0.59 , 2.72	1.25	0.50	0.57 , 2.75

Table 8.6 HR for risk of MI with cumulative use of metformin (top) or sulfonylureas (bottom) compared to diet only.

Models 1-3 are standard pooled logistic regression with varying levels of adjustment for confounders. Model 4 is a MSM with IPTW, and model 5 a MSM with joint IPTW and IPCW. **Basic adjustment** – age, gender, calendar period of diabetes onset, smoking and alcohol **Full baseline adjustment** – Basic adjustment + baseline measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **Time updated adjustment** – Full baseline adjustment plus time-dependent measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **IPTW model** – full baseline adjustment plus stabilised IPTW, where denominator model for IPTW includes all variables in time updated adjustment, and all baseline variables in the numerator. **IPTW and IPCW – IPTW** model plus joint weights for censoring, where censoring models include all baseline and time updated variables in denominator and all baseline variables in numerator.

Figure 8.3: HR curve for cumulative use of metformin (top) or sulfonylureas (bottom) vs diet only, for risk of MI.



95% confidence interval range shown by grey shading. Red Line indicates HR of 1.

8.3.4.2 Stroke

The MSMs estimated an overall increased risk of stroke for metformin use vs diet only, with a HR of 1.29 (1.07-1.55) for the MSM with joint IPTW and IPCW (Table 8.7, left). Standard analysis methods estimated a smaller increased risk of stroke for metformin use vs diet, which was not statistically significant. When examining cumulative medication use, there was no clear association between increasing length of metformin use and risk of stroke, with the HR's remaining within the range of 1.2–1.4, with the exception of 3-6 months, where the HR was 0.96. However, estimates had relatively wide confidence intervals (Table 8.8, top). Plotting the estimated HR curve showed that the model estimated that the risk of stroke was increased even further for greater than 100 months (≈ 9 years) of use. This effect was much stronger in the MSMs than in the standard models (Figure 8.4 top).

For sulfonylureas, there was no evidence of an overall association with stroke for either standard methods or MSMs, with an overall HR of 0.99 (0.71-1.38) (MSM with IPTW and IPCW). When looking at cumulative use, the pattern of results for the MSMs was similar to that for metformin up to 2 years, but from this point onwards the elevated risk was diminished (Table 8.8). When looking at the continuous form of cumulative use, there was some suggestion that the risk of stroke may be decreased for between 4 and 9 years of exposure. This was more slightly prominent in the MSMs compared to the standard analysis methods, but the confidence intervals for all models had similar ranges. (Figure 8.4, bottom).

	Metformin			Sulfonylureas		
	HR	SE	95% CI	HR	SE	95% CI
1 Basic adjustment	1.07	0.06	0.95 , 1.20	1.02	0.12	0.82 , 1.27
2 Full baseline adjustment	1.10	0.07	0.97 , 1.26	1.08	0.13	0.85 , 1.38
3 Time updated adjustment	1.07	0.07	0.93 , 1.22	1.04	0.13	0.81 , 1.32
4 IPTW model	1.27	0.12	1.06 , 1.54	0.99	0.17	0.71 , 1.38
5 IPTW and IPCW model	1.29	0.12	1.07 , 1.55	0.99	0.17	0.71 , 1.38

Table 8.7 HR for risk of stroke with use of metformin (left) or sulfonylureas (right) compared to diet only.

Models 1-3 are standard pooled logistic regression with varying levels of adjustment for confounders. Model 4 is a MSM with IPTW, and model 5 a MSM with joint IPTW and IPCW. **Basic adjustment** – age, gender, calendar period of diabetes onset, smoking and alcohol **Full baseline adjustment** – Basic adjustment + baseline measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **Time updated adjustment** – Full baseline adjustment plus time-dependent measures of: HbA1c, BMI, SBP, history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **IPTW model** – full baseline adjustment plus stabilised IPTW, where denominator model for IPTW includes all variables in time updated adjustment, and all baseline variables in the numerator. **IPTW and IPCW – IPTW** model plus joint weights for censoring, where censoring models include all baseline and time updated variables in denominator and all baseline variables in numerator.

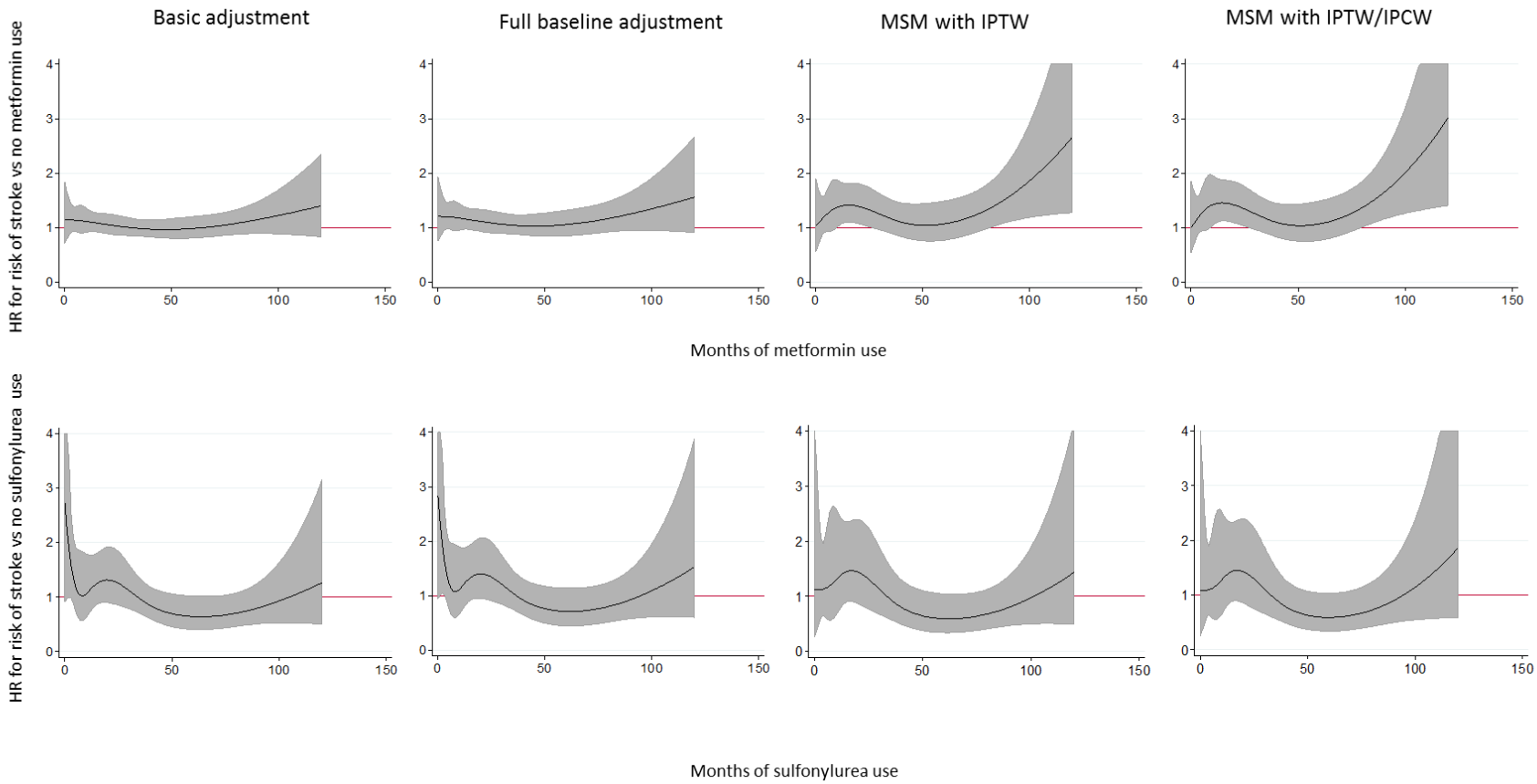
Metformin

	1 Basic Adjustment			2 Baseline Adjustment			3 Time updated adjustment			4 IPTW			5 IPTW and IPCW		
	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI
1- 3 months	1.33	0.20	0.99 , 1.79	1.40	0.21	1.04 , 1.88	1.14	0.18	0.84 , 1.55	1.28	0.24	0.9 , 1.84	1.26	0.23	0.88 , 1.81
3-6 months	0.81	0.16	0.55 , 1.19	0.85	0.17	0.58 , 1.25	0.79	0.16	0.53 , 1.17	0.94	0.29	0.52 , 1.70	0.96	0.31	0.51 , 1.82
6-12 months	1.16	0.15	0.91 , 1.49	1.22	0.16	0.95 , 1.58	1.15	0.15	0.89 , 1.49	1.39	0.27	0.94 , 2.03	1.44	0.29	0.97 , 2.15
1-2 years	1.06	0.11	0.87 , 1.30	1.12	0.12	0.91 , 1.38	1.07	0.12	0.87 , 1.32	1.35	0.22	0.97 , 1.86	1.37	0.23	0.99 , 1.90
2 - 5 years	1.01	0.08	0.86 , 1.19	1.07	0.10	0.89 , 1.28	1.04	0.10	0.86 , 1.24	1.20	0.16	0.92 , 1.57	1.21	0.16	0.93 , 1.58
> 5 years	1.05	0.12	0.84 , 1.31	1.13	0.14	0.88 , 1.44	1.08	0.14	0.85 , 1.39	1.33	0.22	0.96 , 1.84	1.33	0.23	0.95 , 1.86
	Sulfonylureas														
	1 Basic Adjustment			2 Baseline Adjustment			3 Time updated adjustment			4 IPTW			5 IPTW and IPCW		
	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI
1- 3 months	1.74	0.59	0.89 , 3.38	1.81	0.62	0.93 , 3.52	1.19	0.40	0.62 , 2.31	1.12	0.45	0.51 , 2.46	1.08	0.44	0.49 , 2.41
3-6 months	0.86	0.43	0.32 , 2.29	0.90	0.45	0.33 , 2.40	0.83	0.42	0.31 , 2.25	0.81	0.49	0.25 , 2.67	0.72	0.46	0.21 , 2.54
6-12 months	1.32	0.40	0.72 , 2.39	1.38	0.43	0.76 , 2.53	1.33	0.41	0.73 , 2.43	1.65	0.62	0.78 , 3.46	1.61	0.60	0.77 , 3.35
1-2 years	1.31	0.30	0.84 , 2.06	1.39	0.33	0.88 , 2.21	1.35	0.32	0.85 , 2.13	1.13	0.36	0.60 , 2.12	1.11	0.36	0.59 , 2.09
2 - 5 years	0.91	0.17	0.63 , 1.32	0.99	0.19	0.68 , 1.45	0.97	0.19	0.66 , 1.42	1.02	0.27	0.61 , 1.72	1.00	0.27	0.59 , 1.69
> 5 years	0.73	0.17	0.46 , 1.16	0.83	0.20	0.51 , 1.33	0.80	0.19	0.50 , 1.29	0.66	0.18	0.39 , 1.14	0.72	0.22	0.40 , 1.31

Table 8.8 HR for risk of stroke with cumulative use of metformin (top) or sulfonylureas (bottom) compared to diet only.

Models 1-3 are standard pooled logistic regression with varying levels of adjustment for confounders. Model 4 is a MSM with IPTW, and model 5 a MSM wtih joint IPTW and IPCW. **Basic adjustment** – age, gender, calendar period of diabetes onset, smoking and alcohol **Full baseline adjustment** – Basic adjustment + baseline measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **Time updated adjustment** – Full baseline adjustment plus time-dependent measures of: HbA1c, BMI, SBP, history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **IPTW model** – full baseline adjustment plus stabilised IPTW, where denominator model for IPTW includes all variables in time updated adjustment, and all baseline variables in the numerator. **IPTW and IPCW – IPTW** model plus joint weights for censoring, where censoring models include all baseline and time updated variables in denominator and all baseline variables in numerator

Figure 8.4: HR curve for cumulative use of metformin (top) or sulfonylureas (bottom) vs diet only, on risk of stroke.



95% confidence interval range shown by grey shading. Red Line indicates HR of 1. Note: Regression for time updated covariate adjustments would not converge therefore results for this analysis not presented.

8.3.4.3 All- cause mortality

The MSMs estimated overall HRs for ever use of metformin vs diet only on risk of all-cause mortality that were consistent with no association (HR 0.96 (0.86-1.08) for joint IPTW and IPCW, Table 8.9). The standard analysis methods produced similar results. The estimated relative risk of all-cause mortality by cumulative use of metformin is presented in Table 8.10 (top). All three standard analysis models produced very similar results, with estimates and 95% confidence intervals broadly consistent with either no association or a small protective effect for all lengths of exposure. IPTW and joint IPTW and IPCW models were similar to the standard analysis methods for greater than 6 months use of metformin, but had estimates suggestive of a small increased risk for early exposure to metformin, albeit with wide confidence intervals (HR 1.12, 0.77-1.65 for <3 months, HR 1.24, 0.90-1.73 for 3-6 months). The differences between the standard methods and MSMs were more noticeable in the models where cumulative exposure was modelled as a continuous function, though the confidence intervals had broadly similar ranges (Figure 8.5). For the standard models, the general trend was for an early decreased risk that rose to no difference in risk (HR =1), then remained steady as length of exposure increased. In the weighted models, there was an early increased risk which declined relatively quickly to a small decreased risk, then declined further with greater than ≈ 7 years of exposure.

For sulfonylurea use, all models estimated an overall increased risk of all-cause mortality compared to diet only, however the MSM estimated a smaller increased risk compared to the standard analyses, with an overall HR of 1.10 (0.91-1.33) from the MSM with IPTW and IPCW, compared with 1.22 (1.08-1.37) from the unweighted model with full baseline adjustment (Table 8.9). By looking at cumulative use, it became apparent that there was a large increased risk with immediate exposure (1-3 months), with a HR and 95% CI of 2.11 (1.30 – 3.43) for the joint IPTW and IPCW model. This estimated risk declined in a dose dependent manner as length of exposure increased, to an HR of 0.84 (0.62-1.13) for > 5 years exposure. The standard analyses estimated a similar trend but the estimated HR's were generally higher. Though the confidence intervals around the curves were wide, plotting the continuous HR for cumulative exposure to both metformin and sulfonylureas allowed the observed trend from the categorical exposure model to be visualised more clearly (Figure 8.5). All methods clearly suggested an immediate increased risk that was reduced with time over the first 4-5 years, with a smaller immediate risk in both MSMs and the standard analysis with time updated covariate adjustment.

	Metformin			Sulfonylureas		
	HR	SE	95% CI	HR	SE	95% CI
1 Basic adjustment	0.97	0.03	0.91 , 1.04	1.47	0.08	1.33 , 1.63
2 Full baseline adjustment	0.92	0.03	0.85 , 0.99	1.22	0.07	1.08 , 1.37
3 Time updated adjustment	1.01	0.04	0.94 , 1.1	1.37	0.09	1.21 , 1.56
4 IPTW model	0.96	0.06	0.85 , 1.08	1.10	0.11	0.91 , 1.33
5 IPTW and IPCW model	0.96	0.06	0.86 , 1.08	1.10	0.10	0.92 , 1.33

Table 8.9 HR for risk of all-cause mortality with use of metformin (left) or sulfonylureas (right) compared to diet only.

Models 1-3 are standard pooled logistic regression with varying levels of adjustment for confounders. Model 4 is a MSM with IPTW, and model 5 a MSM with joint IPTW and IPCW. **Basic adjustment** – age, gender, calendar period of diabetes onset, smoking and alcohol **Full baseline adjustment** – Basic adjustment + baseline measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **Time updated adjustment** – Full baseline adjustment plus time-dependent measures of: HbA1c, BMI, SBP, history of MI (ever and in previous 3 months), history of stroke (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **IPTW model** – full baseline adjustment plus stabilised IPTW, where denominator model for IPTW includes all variables in time updated adjustment, and all baseline variables in the numerator. **IPTW and IPCW – IPTW** model plus joint weights for censoring, where censoring models include all baseline and time updated variables in denominator and all baseline variables in numerator

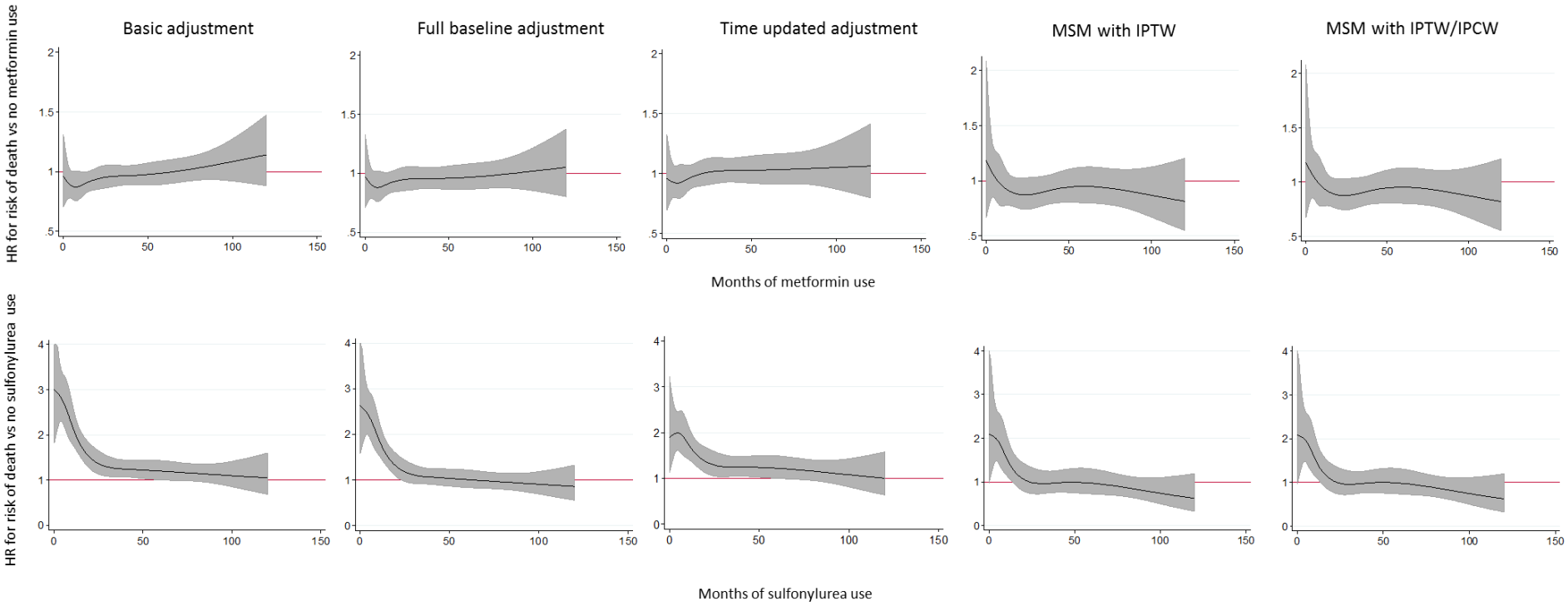
Metformin

	1 Basic Adjust			2 Baseline Adjust			3 Time updated			4 IPTW			5 IPTW and IPCW		
	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI
1- 3 months	0.91	0.10	0.73 , 1.13	0.92	0.10	0.74 , 1.15	0.93	0.11	0.74 , 1.17	1.12	0.22	0.77 , 1.65	1.12	0.22	0.77 , 1.65
3-6 months	1.01	0.11	0.82 , 1.24	1.02	0.11	0.82 , 1.26	1.05	0.12	0.85 , 1.30	1.23	0.21	0.89 , 1.72	1.24	0.21	0.90 , 1.73
6-12 months	0.84	0.07	0.71 , 0.99	0.85	0.07	0.72 , 1.00	0.90	0.08	0.76 , 1.07	0.90	0.12	0.70 , 1.15	0.91	0.12	0.70 , 1.17
1-2 years	0.91	0.06	0.81 , 1.03	0.91	0.06	0.80 , 1.03	0.97	0.06	0.85 , 1.10	0.83	0.08	0.68 , 1.01	0.83	0.08	0.68 , 1.01
2 - 5 years	0.98	0.04	0.90 , 1.07	0.96	0.05	0.88 , 1.06	1.03	0.05	0.93 , 1.14	0.92	0.07	0.79 , 1.07	0.93	0.07	0.80 , 1.08
> 5 years	1.02	0.06	0.92 , 1.14	0.98	0.06	0.87 , 1.11	1.04	0.07	0.91 , 1.19	0.95	0.09	0.78 , 1.15	0.95	0.09	0.78 , 1.15
	Sulfonylureas														
	1 Basic Adjust			2 Baseline Adjust			3 Time updated			4 IPTW			5 IPTW and IPCW		
	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI	HR	SE	95% CI
1- 3 months	2.90	0.47	2.10 , 3.99	2.53	0.42	1.83 , 3.51	1.94	0.34	1.38 , 2.73	2.12	0.52	1.31 , 3.43	2.11	0.52	1.30 , 3.43
3-6 months	2.71	0.46	1.95 , 3.77	2.37	0.40	1.70 , 3.29	1.94	0.34	1.38 , 2.74	1.68	0.42	1.03 , 2.74	1.69	0.41	1.04 , 2.73
6-12 months	2.43	0.31	1.89 , 3.12	2.12	0.27	1.64 , 2.73	1.96	0.26	1.51 , 2.54	1.71	0.29	1.22 , 2.40	1.71	0.29	1.22 , 2.39
1-2 years	1.53	0.18	1.21 , 1.92	1.32	0.16	1.05 , 1.68	1.35	0.17	1.06 , 1.73	1.11	0.22	0.76 , 1.63	1.10	0.21	0.76 , 1.61
2 - 5 years	1.28	0.11	1.09 , 1.50	1.09	0.10	0.91 , 1.29	1.27	0.12	1.05 , 1.52	1.00	0.14	0.77 , 1.31	1.01	0.14	0.78 , 1.32
> 5 years	1.11	0.10	0.92 , 1.33	0.92	0.09	0.76 , 1.13	1.12	0.12	0.91 , 1.38	0.84	0.13	0.62 , 1.13	0.84	0.13	0.62 , 1.13

Table 8.10 HR for risk of all-cause mortality with cumulative use of metformin or sulfonylureas (Top, bottom respectively) compared to diet only.

Models 1-3 are standard pooled logistic regression with varying levels of adjustment for confounders. Model 4 is an MSM with IPTW, and model 5 – joint IPTW and IPCW. **Basic adjustment** – age, gender, calendar period of diabetes onset, smoking and alcohol **Full baseline adjustment** – Basic adjustment + baseline measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **Time updated adjustment** – Full baseline adjustment plus time-dependent measures of HbA1c, BMI, SBP, history of MI (ever and in previous 3 months), history of stroke (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **IPTW model** – full baseline adjustment plus stabilised IPTW, where denominator model for IPTW includes all variables in time updated adjustment, and all baseline variables in the numerator. **IPTW and IPCW – IPTW** model plus joint weights for censoring, where censoring models include all baseline and time updated variables in denominator and all baseline variables in numerator

Figure 8.5: HR curve for cumulative use of metformin (top) or sulfonylureas (bottom) vs diet only, on risk of all-cause mortality.



95% confidence interval range shown by grey shading. Red Line indicates HR of 1.

8.3.4.4 HbA1c

The overall estimates for differences in HbA1c are displayed in Table 8.11. A basic baseline adjustment in an unweighted GEE estimated HbA1c to be higher in patients on both treatments compared to diet only, with estimated absolute differences in HbA1c of 0.71 (0.69-0.72) and 0.95 (0.91-0.98) for metformin and sulfonylureas respectively. Full baseline adjustment resulted in the same direction of difference but of a smaller magnitude. In the weighted GEE's however, use of metformin and sulfonylureas were estimated to reduce HbA1c overall compared to diet only, with absolute differences of -0.22 (-0.24, -0.21) and -0.16 (-0.19, -0.13) for metformin and sulfonylureas respectively, using IPTW and IPCW. Results excluding IPCW were very similar. Covariate specification B produced similar results (see table 23.4, appendix 23).

	Metformin			**Sulfonylureas**		
	Estimated absolute difference in HbA1c (%)	SE	95% CI	Estimated absolute difference in HbA1c (%)	SE	95% CI
1 Basic adjustment	0.71	0.01	0.69 , 0.72	0.95	0.02	0.91 , 0.98
2 Full baseline adjustment	0.18	0.01	0.17 , 0.20	0.26	0.02	0.22 , 0.30
4 IPTW model	-0.25	0.01	-0.27 , -0.23	-0.16	0.01	-0.19 , -0.14
5 IPTW and IPCW model	-0.22	0.01	-0.24 , -0.21	-0.16	0.02	-0.19 , -0.13

Table 8.11 Absolute difference in HbA1c (%) with use of metformin (left) or sulfonylureas (right) compared to diet only.

Models 1-2 are standard pooled logistic regression with varying levels of adjustment for confounders. Model 4 is a MSM with IPTW, and model 5 a MSM with joint IPTW and IPCW. **Basic adjustment** – age, gender, calendar period of diabetes onset, smoking and alcohol **Full baseline adjustment** – Basic adjustment + baseline measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **IPTW model** – full baseline adjustment plus stabilised IPTW, where denominator model for IPTW includes all variables in time updated adjustment, and all baseline variables in the numerator. **IPTW and IPCW – IPTW** model plus joint weights for censoring, where censoring models include all baseline and time updated variables in denominator and all baseline variables in numerator.

When looking at the effect by cumulative exposure, the unweighted models tended to estimate that users of metformin had higher HbA1c compared to those on diet only. The weighted models estimated that within 3 months, HbA1c would be lower on metformin compared to diet only. This reduction was greatest for between 3-6 months with an estimated difference of -0.44% (-0.47%--0.41%). From 6-12 months onwards the estimated differences ranged between -0.17%

and -0.28% (Table 8.12, top). There were only small differences observed between the IPTW and joint IPTW and IPCW estimates.

The two covariate specifications gave very similar results in general, with the exception of the first 3 months of exposure, where the categorical specification (B) estimated that HBA1c would still be higher in those treated with metformin for less than three months. Additionally, from 6-12 months onwards, the estimated differences in HbA1c tended to remain larger (between -0.29% and -0.45%) (Table 23.8, appendix 23).

In the MSM with IPTW and IPCW, the pattern of effect with cumulative use of sulfonylureas was similar to that observed for metformin, though the initial drop within the first 6 months was greater, with an absolute difference of -0.72%, CI (-0.84 , -0.60) for 3-6 months for covariate specification A. Also, the rate at which this initial drop was subsequently attenuated appeared faster, estimating that a patient's HbA1c rose back to that of an untreated patient between 1 and 2 years (Table 8.12, bottom). The MSM excluding IPCW gave very similar results. The differences between model specifications A and B for sulfonylureas showed a similar pattern to that observed for metformin (see appendix 23).

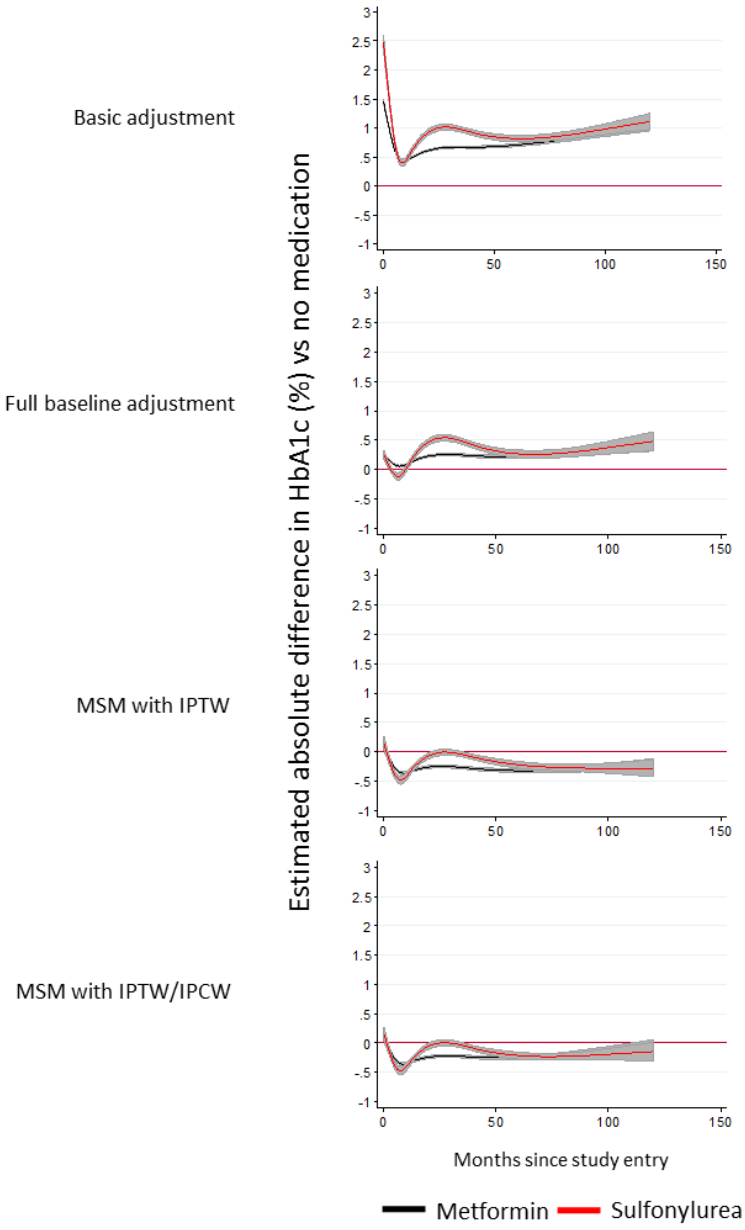
Figure 8.6 displays the estimated absolute difference in HbA1c compared to diet only when cumulative exposure was modelled as a continuous function, clearly demonstrating the changes between standard models and weighted MSMs. The results are broadly consistent with the results in Table 8.12, though here the maximum reduction appears to be somewhere between 10 and 12 months. To visualise how these results translate to predicted HbA1c trajectories, Figure 8.7 shows the estimated trajectories of HbA1c for a patient entering the study with an Hba1c of 7.5% at time 0, if they were to treat continuously with diet only, metformin, or sulfonylureas. The first 5 years of follow up are shown in greater detail on the right side of the figure. In these plots, cumulative use of medication has been modelled with the cubic spline parameterisation. It should be noted that since the plots start at month 1, not all trajectories are expected to start at 7.5% exactly. Analogous results for covariate specification B (categorical parameterisations for covariates) are given in figures 23.4and 23.5, in appendix 23 and show similar findings.

	Metformin											
	1 Basic Adjust			2 Baseline Adjust			4 IPTW			5 IPTW and IPCW		
	Est*	SE	95% CI	Est*	SE	95% CI	Est*	SE	95% CI	Est*	SE	95% CI
< 3 month	1.06	0.01	1.04 , 1.08	0.16	0.01	0.15 , 0.18	-0.10	0.01	-0.12 , -0.07	-0.10	0.00	-0.12 , -0.07
3-6 months	0.50	0.01	0.48 , 0.52	-0.06	0.01	-0.08 , -0.04	-0.44	0.02	-0.47 , -0.41	-0.44	0.04	-0.47 , -0.40
6-12 months	0.59	0.01	0.57 , 0.60	0.20	0.01	0.18 , 0.22	-0.22	0.01	-0.25 , -0.19	-0.21	0.00	-0.24 , -0.18
1-2 years	0.62	0.01	0.60 , 0.64	0.27	0.01	0.25 , 0.29	-0.20	0.01	-0.23 , -0.18	-0.18	0.04	-0.21 , -0.16
2 - 5 years	0.65	0.01	0.63 , 0.67	0.19	0.01	0.17 , 0.22	-0.33	0.01	-0.35 , -0.30	-0.28	0.07	-0.31 , -0.26
> 5 years	0.80	0.02	0.77 , 0.84	0.30	0.02	0.26 , 0.34	-0.26	0.02	-0.30 , -0.22	-0.17	0.01	-0.21 , -0.12
	Sulfonylureas											
	1 Basic Adjust			2 Baseline Adjust			4 IPTW			5 IPTW and IPCW		
	Est*	SE	95% CI	Est*	SE	95% CI	Est*	SE	95% CI	Est*	SE	95% CI
< 3 month	1.69	0.04	1.6 , 1.77	0.05	0.03	-0.01 , 0.11	-0.16	0.04	-0.25 , -0.07	-0.15	0.01	-0.24 , -0.06
3-6 months	0.45	0.04	0.38 , 0.52	-0.48	0.04	-0.56 , -0.40	-0.73	0.06	-0.85 , -0.61	-0.72	0.00	-0.84 , -0.60
6-12 months	0.76	0.03	0.70 , 0.83	0.28	0.03	0.21 , 0.35	-0.19	0.04	-0.27 , -0.11	-0.18	0.00	-0.26 , -0.10
1-2 years	0.92	0.03	0.86 , 0.98	0.55	0.03	0.49 , 0.61	0.05	0.03	-0.02 , 0.11	0.06	0.00	-0.01 , 0.12
2 - 5 years	0.88	0.03	0.83 , 0.93	0.33	0.03	0.27 , 0.38	-0.18	0.02	-0.22 , -0.13	-0.16	0.01	-0.22 , -0.11
> 5 years	0.89	0.03	0.83 , 0.96	0.34	0.04	0.27 , 0.42	-0.22	0.03	-0.28 , -0.16	-0.20	0.00	-0.26 , -0.13

Table 8.12 Absolute difference in HbA1c (%) for cumulative use of metformin or sulfonylureas (Top, bottom respectively) compared to diet only.

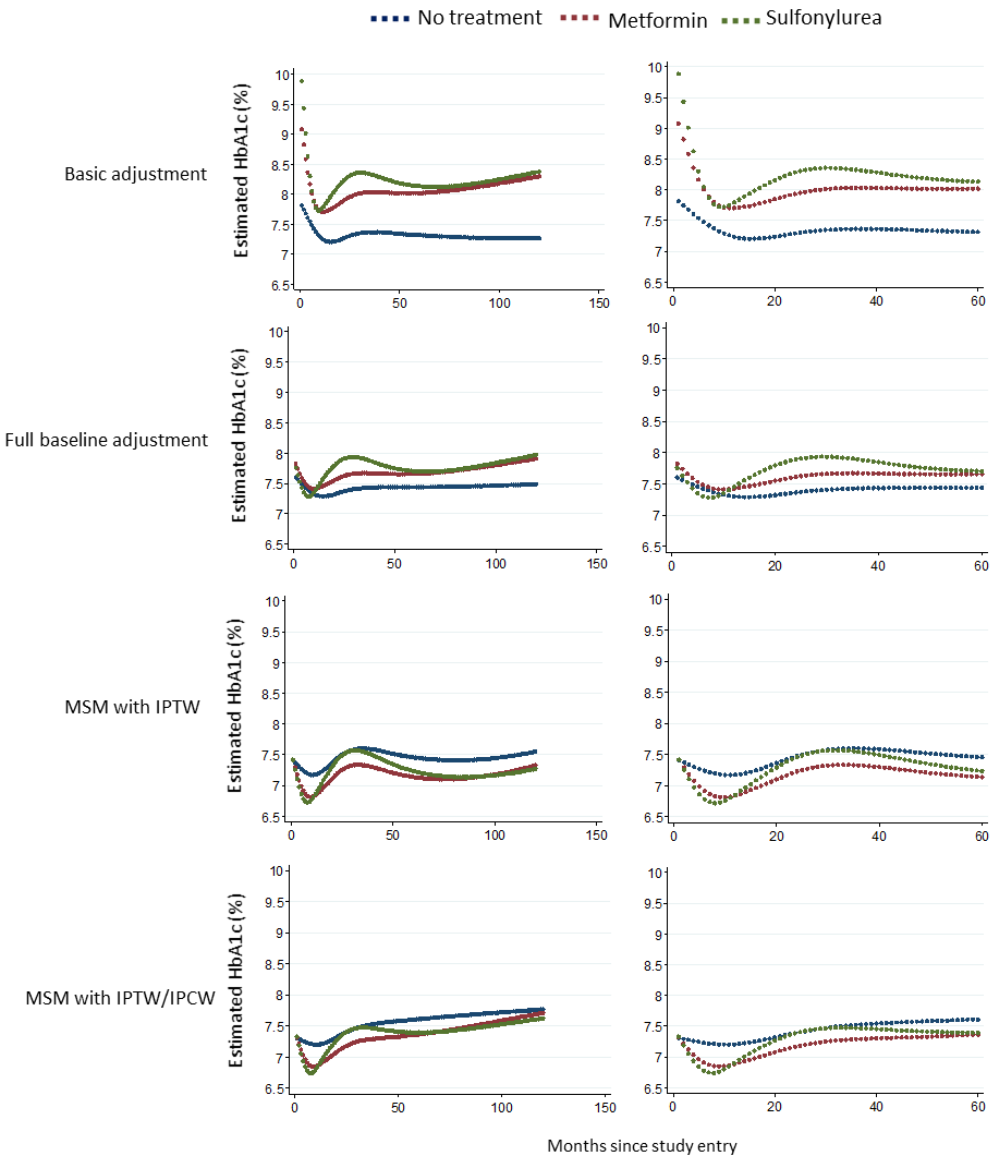
*Est: absolute difference in HbA1c. **Basic adjustment** – age, gender, calendar period of diabetes onset, smoking and alcohol **Full baseline adjustment** – Basic adjustment + baseline measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **IPTW model** – full baseline adjustment plus stabilised IPTW, where denominator model for IPTW includes all variables in time updated adjustment, and all baseline variables in the numerator. **IPTW and IPCW – IPTW** model plus joint weights for censoring, where censoring models include all baseline and time updated variables in denominator and all baseline variables in numerator.

Figure 8.6 Absolute difference in HbA1c (%) compared to diet only with continued use of metformin (black) or sulfonylureas (red).



95% CIs given by grey shading. Straight red line indicated zero difference. **Basic adjustment** – age, gender, calendar period of diabetes onset, smoking and alcohol **Full baseline adjustment** – Basic adjustment + baseline measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **IPTW model** – full baseline adjustment plus stabilised IPTW, where denominator model for IPTW includes all variables in time updated adjustment, and all baseline variables in the numerator. **IPTW and IPCW – IPTW** model plus joint weights for censoring, where censoring models include all baseline and time updated variables in denominator and all baseline variables in numerator

Figure 8.7 Estimated trajectory of HbA1c through time on the three treatment options for full follow up (left) and first 5 years only (right).



Basic adjustment – age, gender, calendar period of diabetes onset, smoking and alcohol **Full baseline adjustment** – Basic adjustment + baseline measures of: HbA1c, BMI, SBP, history of stroke (ever and in previous 3 months), history of MI (ever and in previous 3 months), history of other CVD (ever and in previous three months), ever history of cancer, ever history of CKD, use in the previous 12 months of aspirin, NSAIDS, anti-hypertensive drugs or statins (all drugs entered separately). **IPTW model** – full baseline adjustment plus stabilised IPTW, where denominator model for IPTW includes all variables in time updated adjustment, and all baseline variables in the numerator. **IPTW and IPCW – IPTW** model plus joint weights for censoring, where censoring models include all baseline and time updated variables in denominator and all baseline variables in numerator.

8.3.5 Sensitivity analysis

For most of the additional analyses performed, differences to the primary analysis results were small, and therefore full results are only presented in the main text for some sensitivity analyses. All other results comparing the estimated HRs from the primary analysis with the sensitivity analyses are presented in appendix 24.

8.3.5.1 Age < 65 and BMI ≥ 25 at study entry

Limiting the population to a younger, overweight cohort had no substantial effect on the estimates of effect for MI and stroke. Due to smaller sample sizes and less events, confidence intervals were much wider, particularly for the estimates for use of sulfonylureas. In some instances, no events were observed for some exposure lengths in patients using sulfonylureas so HRs could not be estimated. For all-cause mortality the estimates for metformin use were mostly consistent with the primary analysis. For sulfonylurea use, the estimated HR for risk of all-cause mortality were generally higher compared to the primary analysis in the shorter exposure categories (<3 months, 3-6 months, 6-12 months) though the estimates lacked precision. In the later periods, the changes compared to the primary analysis were minimal (figure 24.1, appendix 24). .

The effect of metformin and sulfonylureas through time on change in HbA1c in this restricted population were estimated to be very similar to the primary analysis, though for 2-5 years and > 5 years use, where for both drugs, the reduction in HbA1c compared to diet only were slightly larger than in the primary analysis (table 24.1, appendix 24).

8.3.5.2 No history of cancer at study entry

Excluding the 178 patients who had incident cancer between time of diabetes diagnosis and study entry made negligible differences to the analysis for all outcomes (table 24.1 and figure 24.2, appendix 24).

8.3.5.3 No previous history of MI/Stroke

Excluding patients with a previous history of MI at study entry made no appreciable differences to estimates of risk of MI for metformin or sulfonylurea use vs diet only. Similarly, excluding patients with no history of stroke made negligible differences to the risk of stroke for metformin users. For risk in sulfonylurea users, there was some suggestion that the estimated HR's were lower compared to the primary analysis, however the confidence intervals for these estimates were still consistent with the primary analysis estimates (figure 24.3, appendix 24).

8.3.5.4 For MI and Stroke, no censoring at the beginning of the interval if an event occurs in the same interval.

Including events in the same interval as a death, increased the number of observed events by 53 for MI and 34 for stroke. Including an event if it occurred in the same interval as a medication change further increased the number of observed events by 17 for MI, and 6 for stroke. However, these additional events made negligible differences to the estimated treatment effects (figure 24.4, appendix 24).

8.3.5.5 Censoring metformin users if they start sulfonylureas and vice-versa

8.3.5.5.1 MI, stroke and all-cause mortality

Use of IPTW (and IPCW to account for the censoring of patients initiating sulfonylureas) resulted in the continued decreased risk of MI (Figure 8.8, top) with use of metformin, albeit with very wide confidence intervals by 10 years of exposure. For stroke, the changes to the primary analysis were minimal, although the estimated increased risk with long term metformin use was much smaller in magnitude. (Figure 8.8, middle). The decreased risk of all-cause mortality with metformin use was more prominent for earlier follow up when patients subsequently starting sulfonylureas were censored, but the continued decrease in risk past 6 years observed in the primary analysis was not apparent, though again confidence intervals by this time were wide (Figure 8.8, bottom).

For use of sulfonylureas, the numbers included in the analysis once those who subsequently initiated metformin were excluded were too low to estimate reliable results for MI and Stroke,

since numbers of events were too low. For greater than four years exposure (approximately), confidence intervals ranged from <0.5 to > 2, with the range spanning 0 to >10 by 7 years, making results uninterpretable. For earlier exposure, results were very similar to the primary analysis but with wider confidence limits. For all-cause mortality, the overall HR curve was similar to that estimated in the primary analysis (Figure 8.9), with two main differences. Firstly, the immediate increased risk was no longer apparent, with instead an initial decreased risk. However, the increased risk over the first two years was still apparent and was estimated to be larger than the primary analysis.

Figure 8.8 Estimated relative risk of MI (top), stroke (middle) and all-cause mortality (bottom) according to months of metformin use from the primary analysis (left) and analysis that censors at sulfonylurea initiation, with updated IPCW (right).

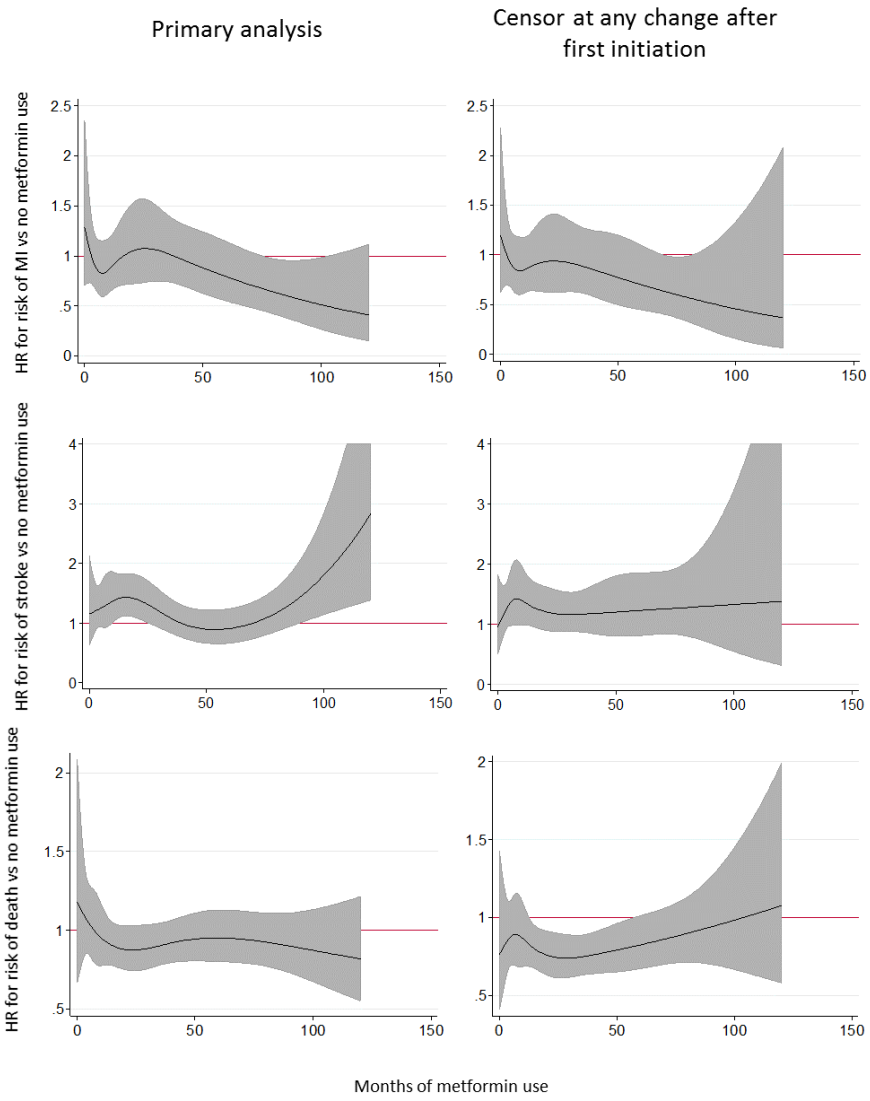
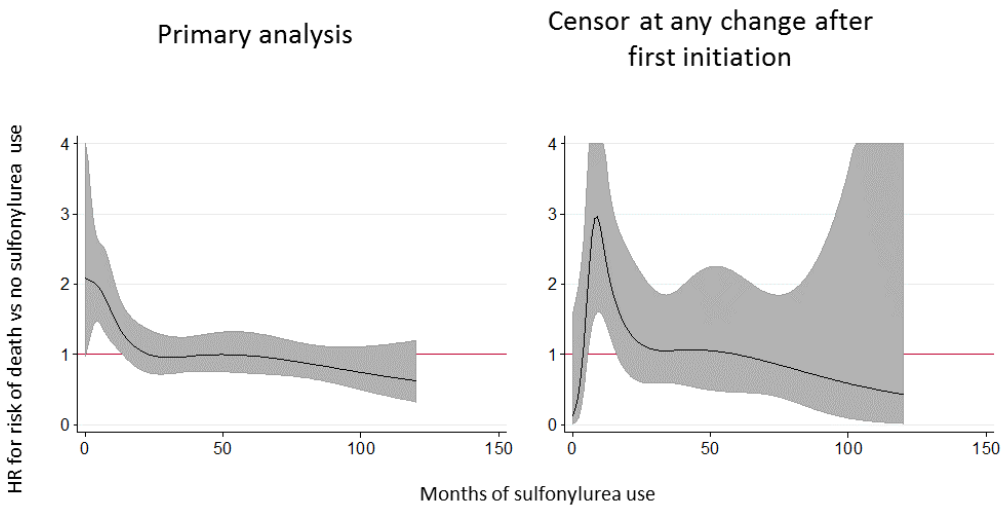


Figure 8.9 Estimated relative risk of all-cause mortality according to months of sulfonylurea use from the primary analysis (left) and analysis that censors at metformin initiation, with updated IPCW (right).



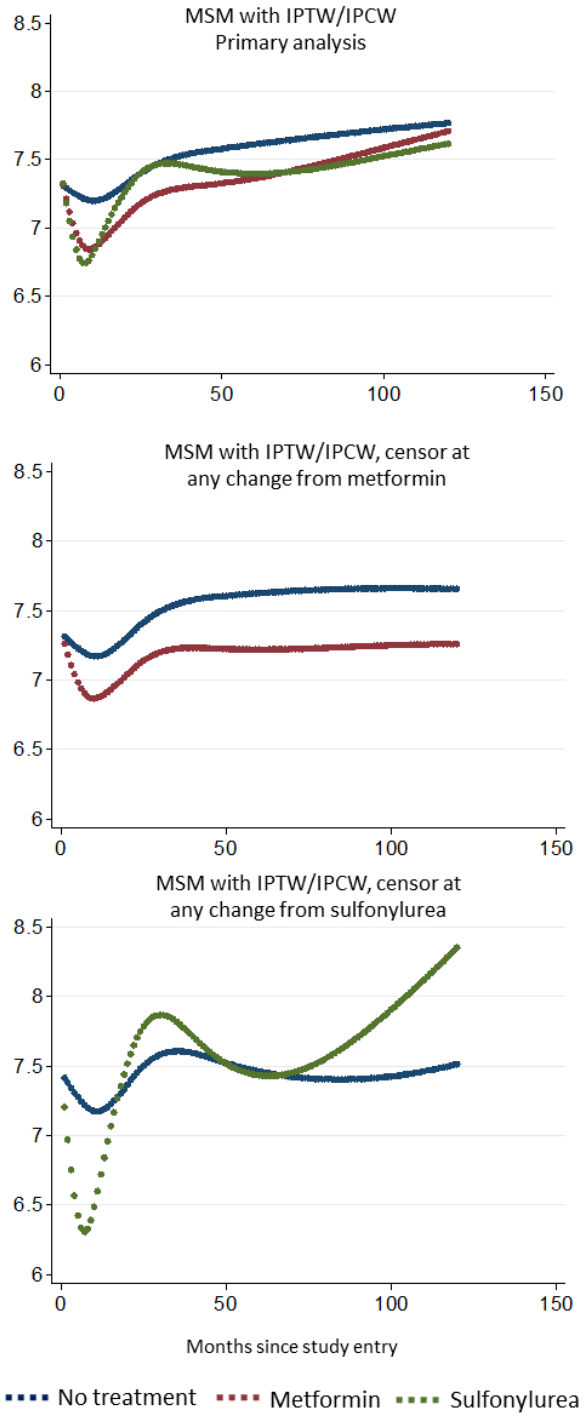
8.3.5.5.2 HbA1c trajectory

For less than 5 years, which is perhaps the most appropriate time interval to consider for this outcome, the estimated trajectories were similar to that of the primary analysis. The main difference for early exposure was that the estimated effect of sulfonylureas on HbA1c reduction was greater compared to the primary analysis (Figure 8.10).

1.1.1.1 *Excluding TIA from definition of stroke*

Excluding TIA from the definition of stroke made only small differences to the estimated risk of stroke for both sulfonylureas and metformin use vs diet only (Figure 24.5, appendix 24). Compared to the primary analysis, for <3 months of metformin use, the estimated slightly HR was lower, though still above one. More noticeably, the increased risk at 1-2 years estimated in the primary analysis disappeared when TIA was removed. The increased risk of stroke with >5 years metformin use was still apparent but was smaller in magnitude. For sulfonylureas, the main changes from the primary analysis were for < 3 months and 1-2 years, where the risk of stroke was estimated to be reduced vs diet only.

Figure 8.10 Estimated HbA1c trajectory through time for treatment with diet only (blue), metformin (red) and sulfonylureas (green) from primary analysis (top), and analyses where patients are censored at sulfonylurea initiation (middle), or at metformin initiaiton (bottom).



8.3.5.6 Addition of extra covariate history for BMI

Addition of a term for BMI two intervals back in the weighting models did not make any appreciable differences to the estimated effects of treatment on any outcomes (table 24.6, appendix 24).

8.4 DISCUSSION

8.4.1 Summary of findings

This analysis applied MSMs with IPTW to estimate the causal effect of metformin and sulfonylureas vs diet only on the risk of MI, stroke, all-cause mortality and HbA1c trajectory through time.

Compared to standard analysis methods, the MSMs estimated a more protective effect of metformin on risk of MI. This was most noticeable with long term exposure. A similar trend was observed for all-cause mortality although confidence intervals were very wide. In contrast, > 8 years exposure to metformin was estimated to increase the risk of stroke. For sulfonylureas, wide confidence intervals made interpretation challenging, and overall there was no evidence of an association between sulfonylurea use and MI or all-cause mortality except for early exposure. There was some suggestion that between 4 and 8 years of sulfonylurea exposure decreased risk of stroke, but this effect was attenuated with longer use. For all outcomes, using both categorical and spline forms for covariates gave similar results, and the differences between MSMs with and without IPCW were negligible.

Models investigating the impact of metformin and sulfonylureas on HbA1c trajectory estimated similar effects for both drugs. The standard analysis methods suggested that HbA1c would be higher (or the same) in patients using metformin or sulfonylureas compared to diet only. In contrast, the MSM estimated that those using metformin and sulfonylureas would have lower HbA1c by three to six months, as expected. This was consistent across all model specifications.

8.4.2 Comparison to the UKPDS study

8.4.2.1 Time-to-event outcomes:

As a randomised trial that compared metformin and sulfonylureas to standard lifestyle intervention, the UK Prospective Diabetes Study (UKPDS) [31] provides a useful comparison for the analyses conducted here . For clarity, because the results of differing covariate specifications were very similar, throughout this section, comparison will be made with the results of the spline models only (specification A).

The UKPDS estimated an overall HR for metformin vs diet on risk of MI of 0.61 (0.41-0.89), based on a median follow up of 10.7 years. The equivalent joint IPTW/IPCW estimate for this analysis, with a median follow up time of 4 years, was 0.93 (0.73-1.18).For greater than 5 years of exposure, the MSM estimated a HR more consistent with that of the UKPDS (0.65 (0.42-1.01)). The crude event rate was also much higher in the UKPDS study, with an event rate of 11 per 1000 person years in the metformin group, and 18 per 1000 person years in the conventional therapy group, compared to an overall rate of 3.9 per 1000 person years in the present analysis. For all-cause mortality, the UKPDS estimated a protective effect of metformin with an overall HR of 0.64 (0.45 – 0.91), compared to 0.96 (0.86-1.08) estimated in this analysis. The moderate protective effect observed here for long term exposure (HR approximately 0.80 at 10 years but with wide CI's) was also inconsistent with the UKPDS study overall. The overall rate of stroke observed in the UKPDS was more consistent with the rates observed here. However as with MI and all-cause mortality, the overall HR estimates for metformin vs diet for risk of stroke were not consistent (0.59 (0.29 – 1.18) for UKPDS vs 1.29 (1.07-1.54) for this analysis.

The UKPDS presented direct comparisons between sulfonylureas and diet by two different classes of sulfonylureas, though results for the two classes were similar. For simplicity the comparison here is made to the UKPDS results for chloropropamide. The estimated overall HR's for MI, stroke and all-cause mortality for chloropropamide vs diet only were 0.87 (0.68-1.12) ,1.01 (0.65-1.58) and 1.02 (0.82-1.27) respectively. The analogous estimates obtained using MSMs (joint IPTW/IPCW) in our analyses were 1.01 (0.68 – 1.52) 0.99 (0.71-1.38) and 1.16 (0.97-1.40). With the exception of all-cause mortality, these results were more consistent with the UKPDS than those for metformin were.

There are several possible reasons for the observed differences in estimates between UKPDS and the analyses conduced in this chapter. Firstly, it is likely that there is residual confounding in the analyses performed here, in contrast with the randomised comparison made by the

UKPDS. The general issue of residual confounding in terms of possible explanation for results will be discussed in 8.4.3 and 8.4.4.1. Potential explanations specific to UKPDS are the issues of differing population, follow up, and time periods between the two studies.

The UKPDS study began in the 1970's, and the sub study on metformin vs diet alone was conducted in overweight patients with newly diagnosed type 2 diabetes, aged under 65 at the time of randomisation. Although the younger age group would have a reduced risk of MI, stroke and death compared to the population studied here, the restriction to overweight patients may increase their risk. The sensitivity analysis performed in patients aged <65 and with a BMI $\geq$.25 at study entry did not provide results that were more similar to those obtained in the UKPDS study. The estimated effect of sulfonylureas on all-cause mortality was more harmful in this restricted population, making it less consistent with UKPDS. Young overweight patients were less likely to be prescribed a sulfonylurea in our data, and so this result may be explained by the possibility that the indication for the prescription was more strongly related to risk of death in this sub-population, for reasons that were not fully accounted for by the included covariates. Overall however, the restriction in sample size led to a small number of observed events and as such wide confidence intervals, so it was not possible to confidently determine whether differences in age and BMI distribution between the two populations could explain the observed differences in effect estimates.

Rates of stroke, MI and death are not constant across diabetes duration, and so the lower observed event rates (particularly for MI) in this analysis compared to the UKPDS, could also be explained by shorter follow up. This is supported by the fact that the event rates in the CPRD population dropped substantially when the age was restricted to <65 years at study entry. With a median follow up of around 4 years, these patients may not have had sufficient follow up to observe cardiac events, or the effect of metformin on such events.

Finally, the calendar time period during which the two studies were based were different. UKPDS began follow up in the 1970s, whereas the analysis here was from 2000 onwards. As discussed by Ferrannini [133] statins were not licenced until the 1990s, whereas nearly 50% the population studied here were on statins at study entry. This may partly explain the difference in event rates. It is also possible that any protective effect of metformin on these outcomes was diluted due to high prevalence of statin use.

8.4.2.2 HbA1c trajectory

The UKPDS also examined change in Hba1c during the study period. The authors observed around a 1% drop in HbA1c for metformin and sulfonylurea users in the first year to 18 months of follow up, before it began to rise to a point still below the baseline value at three years. In contrast, those on a lifestyle intervention saw a near linear increase in HbA1c of about 0.5% over the same period.

The results from the MSMs obtained here were broadly similar. The absolute changes in Hba1c varied slightly depending on model specification but in general, both metformin and sulfonylureas were estimated to reduce HbA1c by around 0.5-0.8% over the first 12 months, followed by an increase back up to the levels of patients on diet only by around 3 years, which in this case was slightly higher than the baseline value (when set at 7.5%). When comparing to diet only, the absolute differences observed here were slightly smaller than those in UKPDS since the diet group were also observed to have a small reduction in HbA1c in the first year. After 3 years of follow up, although there was still an overall trend for metformin and sulfonylurea users to have lower in HbA1c compared to diet only, the magnitude of these differences became less consistent with what was observed for this time period in the UKPDS. The results of the sensitivity analysis restricting to a younger overweight population gave similar results to the primary analysis, so remained consistent with UKPDS findings. The smaller initial absolute drop in HbA1c in the present analysis compared to UKPDS may be explained by adherence to medication. In UKPDS, as patients were enrolled in a trial, they may have been more likely to take their medication regularly. However, in the present study, no minimum adherence was required. Other studies that have examined adherence using routinely collected data have estimated differences in HbA1c over the first 6-12 months to be less in patients with sub optimal adherence. For example, in a study by Nichols et al [223], 50-79% adherence to metformin (based on days covered by prescriptions) was estimated to reduce HbA1c by 0.45% compared to 0% adherence, which, assuming diet only is similar to 0% adherence, is consistent with the change observed in our analyses.

8.4.3 Comparison between standard methods and MSMs

In contrast to the analysis of metformin and cancer, the difference in effect estimates between standard analysis methods and the MSMs were more noticeable for the CVD, morality, and HbA1c outcomes considered in this chapter. The most substantial difference was between

models for longitudinal HbA1c. Prior values of the HbA1c are themselves a time-dependent confounder, and therefore as was observed, a standard adjustment for baseline confounders (baseline HbA1c in particular) will estimate a relative difference in HbA1c between those treated and untreated which makes treatment look ineffective or harmful, because those treated after baseline have higher HbA1c. However, the MSM estimated that that both metformin and sulfonylureas reduced HbA1c compared to diet only, most effectively over the first 12 months of exposure. While acknowledging that the actual effect estimates of the MSM may still lack precision, or have small residual bias, a large difference in effect estimates in the expected direction between the two analysis methods serves as good evidence that use of MSMs in the diabetes context can estimate known effects in the presence of strong time-dependent confounding.

The results for risk of MI with prolonged metformin use, although not as prominent, also add weight to this conclusion. The analyses in section 8.3.2, suggested that higher HbA1c in the previous interval was predictive of higher risk of MI. With high HbA1c also being a strong predictor for metformin initiation, the standard analysis would be expected to bias the effect estimate in the direction of a more harmful effect for both drugs. Compared to all 3 standard analysis methods, the estimates from the MSMs were more indicative of protective effect of metformin, suggesting that the MSM was performing as would be expected. In contrast, there were less clear differences between models for the effect of sulfonylurea use on MI. Other time-dependent risk factors for MI such as SBP, BMI, and statin use that were also included may also have affected the difference between the two models. In particular, obesity (as measured here by BMI) is inversely associated with sulfonylurea use and positively associated with risk of MI [224]. The opposite direction of confounding for BMI and HbA1c could explain the lack of clear effect of the weighting process on these estimates.

For stroke, the changes between the MSM and standard analyses were not in the expected direction. Existing literature would suggest that higher values of both BMI and Hba1c would increase the risk of stroke [225-227]. Therefore, for both metformin and sulfonylureas, the direction of time-dependent confounding would be expected to be the same as for MI. However, in general the same patterns were not observed and the HRs for metformin suggested a more harmful effect in the MSM compared to the standard analysis. It is possible that the changes are random, since confidence intervals were relatively wide. The unexpected direction of confounding could also be due to the use of an ITT principle, in that by the time that the increased risk of stroke became most prominent, patients appearing to remain on metformin had actually intensified to joint therapy with sulfonylureas and so had an increased risk of stroke

compared to patients still treating with diet only. The observed reduction in the magnitude of the increased stroke risk for long term metformin users when an as treated approach was taken (see 8.3.5.5) supports this possibility, but a small increased risk was still observed in this analysis, albeit with wide confidence intervals. It is also possible that the definition of stroke was not precise enough, since different kinds of stroke have slightly different risk factors. This limitation will be discussed in more detail in 8.4.5.2.

In contrast to non-fatal events such as MI and stroke, it is possible that being at high risk of death in the early stages of diabetes could reduce the likelihood of initiating treatment, even if the patient had a high HbA1c. In a standard analysis, this may result in a protective effect of short term treatment. The standard analysis estimated that there was a small decreased risk of death with metformin use for 0-3 and 3-6 months of exposure, while the MSM suggested a small increased risk. This change in effect is consistent with the removal of confounding due to extreme frailty, though the harmful effect of metformin (albeit with wide CIs) estimated by the MSM may suggest further residual confounding in the other direction. The same direction of confounding for early use may also be expected for sulfonylureas, but this was not observed. The standard analysis estimated a substantial harmful effect of early sulfonylurea use, which was reduced but not removed by the IPTW. This could suggest instead, that those at greatest risk of mortality in the short term are treated with sulfonylureas rather than metformin. Such an observation may be explained by confounding by unmeasured factors that indicate severe illness, one of which may be severe chronic kidney disease. Due to minimal numbers of patients with stage 4 or 5 CKD initiating metformin, presence of chronic kidney disease was dichotomised into none, or stage 3, 4 and 5 combined. Therefore, patients with CKD at time of treatment initiation who started metformin over sulfonylureas were more likely to have stage 3 CKD, and may be less likely to die than patients initiating sulfonylureas that have stage 3, 4 or 5 CKD. A simple descriptive investigation (see appendix 13) showed that 1.25% of person months in which a sulfonylurea was initiated were classified as stage 4 or 5 CKD. This is in contrast to 0.33% of person time prior to any treatment initiation (i.e. 0.33% of all “diet only” person time). Additionally, although the actual numbers were small, 0.007 deaths per person month (6/842) exposed to sulfonylureas with any CKD were observed in the first 12 months of follow up, in contrast to 0.003 per person month on diet only with any CKD (125/36169). Two of the 6 deaths observed among users of sulfonylureas were in stage 4 or 5 CKD, compared with only 11 of the 125 deaths among those on diet only. These findings support the possibility that later stage CKD may partly explain the observed early risk of death with sulfonylurea use. In addition, although severity of CKD may also be associated with an increased risk of MI and stroke, this may be

accounted for by other covariates in the weighting models such as use of statins, use of anti-hypertensive medications, and history of CVD, meaning that the inability to perfectly adjust for CKD stage is less likely to be problematic for the other outcomes. It is possible that the results from observational studies [125] discussed in chapter 3 that found a similar harmful effect of sulfonylureas when compared to metformin could have a similar explanation.

8.4.4 Validity of assumptions

8.4.4.1 No unmeasured confounding

It is possible that the MSMs fitted here were affected by some residual or unmeasured confounding, particularly for all-cause mortality (as discussed in the previous paragraph). In addition to the possibilities discussed above, other unmeasured variables discussed in the previous chapter, such as exercise and smoking quantity, are known to be associated with MI, Stroke and mortality [228, 229]. Additionally, cholesterol is also a strong risk factor for cardiovascular disease and may also influence a GPs decision to initiate treatment, but was not included as a covariate in these analyses. Although measures of HDL and LDL cholesterol are available in the CPRD, there was concern over further reduction in numbers due to requiring complete data before treatment initiation. Omission of variables such as diet, exercise and cholesterol levels may have also affected how well the censoring weights were able to adjust for treatment switching in the sensitivity analysis where patients were censored at their first change after first line treatment, since all these factors could plausibly influence how quickly a patient needs to intensify. Another issue is the lack of detail in the adjustment for some of the measures of concomitant medication, such as the dosage or length of exposure to statins.

8.4.4.2 Model misspecification

To see whether different specifications of covariates in the models may affect the estimated treatment effects, both spline forms and categorisations of model parameters were used in the treatment, censoring and outcome models. There were no large differences between categorical and spline models, but a few small differences were observed, which may suggest issues with model misspecification. For example, in general, lower HRs were estimated for both metformin and sulfonylureas in the categorical models for all-cause mortality (see appendix 23, tables 23.3 and 23.7), which, for example, would be consistent with residual confounding by

age, since younger patients are more likely to be treated and older patients are at higher risk of death.

There may also have been some issues of misspecification with the spline parameterisations, since the same splines for baseline covariates and time were used in the model for the weights and outcome model, and these parameterisations were based on the association with treatment. As explained in 8.2.5.2.2, it was felt that this would be preferable for ease of data management. The issues with this have been previously discussed (see section 7.4.3.3). As explained in 8.2.5.2.2, with the exception of baseline BMI for all-cause mortality, which was re-parameterised accordingly; a simple check suggested that this approach was unlikely to have caused serious model misspecification (see appendix 21).

The models for the weights assume that full covariate history is modelled. In this analysis, baseline, and the most current value of time updated covariates were included. For HbA1c, a further variable to represent the value two intervals back was also included after observing that it was also predictive of outcome (see section 8.3.2). However, no further variables for covariate history were considered. Although the reason for this was due to concerns over collinearity due to relatively infrequent updating of covariate values, it is acknowledged that by only considering more recent history of covariates, that full covariate history may not have been perfectly modelled, and as such any time-dependent confounding not perfectly removed.

It was observed that there were minimal differences between the MSM with IPTW and IPCW, and IPTW only. For purely administrative censoring, this seems reasonable since there is no clear reason why this could be associated with risk of outcome. However, one reason for censoring was if treatment was intensified to something other than metformin or a sulfonylurea. This might be important since more severe diabetes would likely be predictive of all outcomes, and of treatment intensification. One possibility is that the censoring model was poorly specified. For example, although current treatment was included in the censoring model, time on treatment was not. However, since the number of patients censored for this reason was relatively small, the impact of such misspecification is likely to be minimal. As shown in chapter 6 (see tables 6.1), 86% of patients who initiated metformin as a first line therapy intensified to sulfonylureas, and 70% of patients who initiated sulfonylureas intensified to metformin as their second line treatment. Therefore, the censoring at other treatment was predominantly reflecting censoring at third line intensification, which was only observed during follow up to all-cause mortality for 15% of the population (see section 6.3.1).

Finally, the MSM with ITPW assumes that the full treatment history is modelled in the outcome model. To thoroughly investigate the effect of treatment history, both current treatment and cumulative treatment were considered in separate models. Results suggested that current treatment alone was not sufficient to fully describe treatment history in terms of the effect of treatment, but since other studies e.g. UKPDS present single HR's for treatment effect, it was still felt useful to estimate the effect of a binary current treatment effect for comparison.

8.4.4.3 Positivity

The general issues around the positivity assumption (as discussed in 7.4.3.4) also apply here. More specific to this analysis, it was noticeable that the initial distribution of the weights was much improved compared to the previous chapter, in that the maximum weights, and as such the mean of the weights were smaller. It is possible that this is due to the fitting of separate treatment models by calendar period, and because only patients diagnosed after the year 2000 were included. However, there were still some extreme weights observed and so truncation was still necessary. It was also noticeable that the categorical model specification had a lower mean for the untruncated stabilised IPTW, suggesting that this coarser categorisation was further reducing issues with positivity, but the 99th percentiles were similar between specifications. Once the truncation was applied therefore, the differences between the weights was minimal, which could explain the small differences observed between covariate specifications, and it is acknowledged that such a truncation may have removed the benefit of using the categorical specification to obtain less extreme weights

8.4.5 Other limitations

8.4.5.1 Visit frequency

The implications of GP visit frequency being potentially dependent upon underlying health, for weight estimation and for the overall interpretation of studies using EHR data, have been discussed in detail in the previous chapter (see section 7.4.4), and the issue may also have affected this analysis. It is possible that frail patients at risk of death may be less likely to visit their GP for routine care as they are too weak to attend. This may also prevent them from receiving treatment, meaning that the issue of non-attendance may be more problematic for the outcome of death. As an example, patients diagnosed with diabetes who subsequently

become ill may then be unable to visit the GP for diabetes related care. If so, they would have no updated records to accurately predict risk of death, and would also be unable to receive diabetes treatment. Equally, frail patients may enter hospital care so their prescribing may move away from the GP (described by Suissa as immeasurable time bias [230]). It is conceivable therefore, that that the trend towards a protective effect of long term use of metformin and sulfonylureas on all-cause mortality could be residual confounding driven by patients who have a diagnosis for diabetes but then become unable to attend/enter hospital care and subsequently die.

The results of a repeated measures GEE may also be affected by visit frequency issues. Using an unbalanced repeated measures model for the MSM means that the HbA1c trajectory is based only on patients who are attending the GP. However, the weighting uses LOCF for HbA1c values. Therefore as with the other outcomes, it is possible that the weighting does not appropriately balance the effect of the time-dependent covariates on HbA1c trajectory between the different treatment groups. Non-attending patients may have worsening HbA1c because their nonattendance is reflective of insufficient diabetes management. By not visiting, they cannot be treated, and therefore by the time they do visit the GP, remaining untreated will estimate higher HbA1c. In contrast, regular attendance will likely pick up changes in HbA1c that indicate the need for treatment, which will result in better HbA1c control in those treated. This issue would also affect the standard analyses, however it could be amplified in the weighted MSM if, for example, patients not attending are more likely to be upweighted based on the covariate values being carried forward. This will depend on where non-attending subjects start in terms of the HbA1c scale and other covariates, and whether the assumed trajectory of HbA1c while not attending is correct, making the direction of bias difficult to predict. Having said this, it is reassuring that the change in estimates with the MSMs compared to the standard analysis is consistent with what would be expected in the presence of time-dependent confounding and that the estimated trajectories in the MSM were broadly consistent with what has been previously observed in trials, which suggests that the issue of potentially differing visit schedules has not severely affected this analysis.

8.4.5.2 Outcome definitions

Identification of patients experiencing the outcomes of interest was based solely on the information provided in the primary care data files from CPRD. A study by Herrett et al [231] showed that this may underestimate the incidence of MI in the general population by

approximately 25% compared to using linkage with disease registries and Hospital Episode Statistics. This is also likely to be true for stroke, and is an acknowledged limitation of this study in terms of reduced precision. However, there is no clear reasoning to suggest the underestimation would be systematically different between treatment groups, so this is unlikely to have caused serious bias. Secondly, in a diabetic population, patients are more likely to have contact with their GP and therefore get the events recorded, so the underestimation may be lower.

The rate of MI in particular was observed to be much lower in the present study than in the UKPDS study. As already discussed in 8.5.2, this may be due to the effects of statin use and shorter follow up. Herrett et al [231] estimated the crude incidence of MI in the general population to be 1.87 per 1000 person years when estimated using CPRD primary care data only, in the period from 2003-2009. A second study by Shah et al [18] aimed to estimate the relative risk of various cardiovascular events between diabetes and non-diabetes, finding an increased incidence of non-fatal MI (HR 1.54 (1.42 – 1.67)) in patients with diabetes compared to patients without diabetes. A different study estimated this relative risk to be closer to 2 for both fatal and non-fatal MI in subjects with no previous history of MI [232]. Considering these results together would give a rough incidence of MI in a diabetic population of 2.87 per 1000 person years (assuming a relative risk of 1.54), to 3.4 per 1000 person years (assuming a relative risk of 2). The estimate of crude incidence of both fatal and non-fatal MI observed in the present study of 3.9 (3.6 – 4.1) per 1000 person years is roughly consistent with these studies, when also accounting for the fact that patients with previous history of MI were included, which will increase the incidence.

Another limitation of these analyses was the broad definition used to define stroke. Inclusion of TIA may not have been appropriate since it is a less severe event, which may be poorly recorded and potentially misclassified in primary care records. Sensitivity analysis showed that exclusion of TIA as a stroke event did not materially change the effect of either metformin or sulfonylureas on overall stroke risk, except for 1-2 years of exposure, where the estimated HR was reduced for both drugs.

The definition of stroke also combined ischaemic and haemorrhagic stroke. The same study by Shah previously mentioned [18] also looked at the risk of different types of stroke between patients with and without diabetes. They found that the incidence of ischaemic or unspecified stroke was increased in patients with diabetes, but that risk of subarachnoid haemorrhage was lower in patients with diabetes (HR 0.48, (0.26-0.89)). The authors also found that the risk of intracerebral haemorrhage was similar between patients with and without diabetes, particularly

for those aged less than 75. It is possible that these differences between stroke types may also be relevant for comparisons within patients with diabetes on different treatments, for example if use of metformin alters the risk to that more reflective of a patient without diabetes. The analyses of other outcomes are less likely to be affected by the combination of different types of stroke, since in general, history of any CVD event increases the risk of a future event.

8.4.5.3 Follow up time

As in the previous chapter, the relatively short follow up time and restriction to newly diagnosed diabetes may have limited the possibility of capturing the full effects of treatment on outcomes such as all-cause mortality, MI and Stroke. Although follow up was lengthened by allowing patients to switch between metformin and sulfonylureas, precision of the estimated treatment effects for long term exposure was still low. This limited the ability to draw firm conclusions about the effects of treatment through time. For longitudinal HbA1c, the effect of treatment is more short term, so insufficient follow up was less of a limitation. It should be noted that the interpretation of the estimated treatment effect was altered by allowing switches between the two treatments of interest. The primary analysis estimated the effect of being prescribed metformin (or a sulfonylurea) as a first line therapy compared to remaining untreated. This differs to the sensitivity analysis that censored at switch from metformin to sulfonylureas or vice versa. Here, the analysis estimated the effect of being prescribed metformin (or a sulfonylurea) as a first line therapy compared to diet only, where patients remain on metformin (or a sulfonylurea) only for their entire follow up.

8.5 CHAPTER SUMMARY

MSMs with IPTW were applied to estimate the causal effect of metformin and sulfonylurea use vs diet only on the risk of MI, stroke, all-cause mortality and HbA1c trajectory through time. The primary analyses indicated that long term use of metformin may be protective against incidence of MI and all-cause mortality, though confidence intervals were wide. In subsequent sensitivity analyses where patients were censored if they switched from metformin to a sulfonylurea, the estimated long term protective effect of metformin on mortality was reduced, though overall, results remained consistent with a beneficial effect. There was no strong suggestion that sulfonylurea use affected risk of MI in either direction. As has been observed in previous studies, standard analysis methods estimated a large increased risk of all –cause mortality with

sulfonylurea use, which was reduced, but not entirely removed with use of the MSMs. It is probable that residual confounding explains this observation, though the possibility of a harmful acute effect cannot be excluded. There was also some suggestion that long term metformin use increased the risk of stroke. This effect was somewhat reduced with removal of TIA from the outcome definition, and when time on dual treatment with sulfonylureas was excluded. Further work is required to understand the possible reasons for this.

In contrast with the previous chapter on metformin and cancer, use of the MSMs to control for time-dependent confounding appeared more effective for analysing the outcomes presented in this chapter. This was particularly noticeable for modelling longitudinal HbA1c, where the use of the MSM recovered the expected findings that both metformin and sulfonylureas reduced HbA1c compared to no treatment. The large difference in effect estimates in the expected direction between the standard and causal analysis methods for examining HbA1c trajectory serves as good evidence that use of MSMs in the diabetes context can recover plausible drug effects in the presence of strong time-dependent confounding. For outcomes where the causal pathway between treatment and outcome are predominantly indirect and long term, residual confounding is likely to be an issue, and more long term data than was available here may be needed to draw confident conclusions.

9 DYNAMIC MARGINAL STRUCTURAL MODELS TO COMPARE HBA1C INITIATION THRESHOLDS FOR FIRST LINE TYPE 2 DIABETES TREATMENTS

9.1 INTRODUCTION

Current treatment guidelines in the UK suggest initiation of first line diabetes therapy if HbA1c cannot be controlled at under 6.5% with a diet and lifestyle intervention [13]. However, the review of literature in section 3.3 of this thesis suggested that there was no clear evidence supporting this specific threshold. Further to this, a direct comparison of alternative first line initiation strategies based on HbA1c monitoring could not be found within the search of published literature. As described in section 4.5, dynamic marginal structural models allow comparisons of alternative dynamic strategies in observational data sets, provided there are no unmeasured confounders of the association between strategy compliance and outcome.

In the previous chapter, use of weighted MSMs for comparison of static first line treatment strategies showed that initiation of metformin or a sulfonylurea was associated with subsequent lowering of HbA1c during follow up in comparison to diet only. There was also some suggestion that metformin and sulfonylureas may affect risk of MI, stroke and all-cause mortality, but as discussed (see section 8.4), there were limitations to these analyses and the estimated effect of treatment on some outcomes lacked precision. In addition, the previous analyses did not examine dynamic strategies and questions about when to initiate treatment. Given that current treatment guidelines are predominantly based around HbA1c thresholds, comparison of dynamic strategies are highly clinically relevant. Comparison of dynamic strategies can also overcome issues of positivity violations if the strategies to be compared are observed to have reasonable compliance.

The following analyses aim to compare different thresholds of HbA1c for initiation of first line treatment, with regard to a) the time taken to reach target HbA1c, and b) the relative risks of MI, stroke and all-cause mortality. Results of such analyses may have implications for management and monitoring of early stage T2DM.

9.2 METHODS

9.2.1 Study population

The study population was the broadly the same as in the previous chapter (see sections 8.2.1 and 8.3.1). Briefly, it consisted of patients with newly diagnosed type 2 diabetes from the CPRD, diagnosed from the year 2000 onwards. As before, patients entered the study at the earliest time after date of diabetes diagnosis that they obtained complete data (see section 4.4.2.1) and were considered at risk in the analysis for a maximum of 10 years until the outcome of interest, death, transfer out, initiation of any treatment other than metformin or sulfonylureas, or end of data collection (January 2016). A key difference was that for the dynamic MSMs, only patients who were treatment free at study entry (n=49,750) were included. This was because when using unstablised weights (as described in 4.5.3.3), it is not necessary to adjust for baseline confounders in the dynamic MSM, as this is done by the weighting. If patients treated from study entry were included, then the baseline covariate adjustment in the MSM (see section 7.2.5) would have been necessary, therefore it was decided for simplicity that they should be excluded.

9.2.2 Defining strategies for comparison

For this analysis, the dynamic strategies to be compared were of the form **"treat with metformin or a sulfonylurea when HbA1c first rises above x%"** with variable x. Using the same ITT style approach to the primary analysis in chapter 8, switches between metformin and sulfonylureas after initial initiation were allowed, meaning that the strategy definition could be more precisely defined as "treat with metformin or a sulfonylurea when HbA1c first rises above x% and continue with either or both medications". The choice to allow treatment with either metformin or a sulfonylurea was made to allow greater numbers to be compliant to the different strategies. In doing this, we made the assumption that both drugs have the same effect on the outcome of interest. Since the results of the previous chapter suggested this may not be the case for CV outcomes and mortality, sensitivity analyses were performed to evaluate metformin only initiation strategies (see 9.2.7).

As explained previously (see section 4.5.4.1), the values for initiation thresholds that can be investigated (denoted x) depend upon what is actually observed. If all GPs abide strictly to the current guideline of 6.5%, then there will be only one value for x with which patients are compliant. It was decided to aim to look at strategies within the range of 6.5-10%, as this was

considered a clinically relevant range, with initial descriptive analysis used to determine whether the observed data would allow this range of thresholds to be investigated. For simplicity, compliance with a strategy was based only on first line initiation. If a patient initiated treatment in line with the strategy under consideration, they were considered compliant for the rest of their time at risk. As in previous chapters, censoring at initiation of medications other than metformin or sulfonylureas was treated as loss to follow up rather than subsequent noncompliance. For the thresholds of 6.5, 7, 8, 9 and 10%, the number of patients compliant with each strategy through time (i.e. still compliant and still at risk at a given time), and the overall proportion of patients compliant to each strategy up to and including the point of treatment initiation, referred to as "overall compliance" were calculated. Any strategies with low overall compliance, as defined by <10%, were removed from the set of strategies to be investigated further. Ideally, many values for x over a particular range would have been used to model the strategies as a continuous function in order to identify optimal initiation thresholds. However, due to the size of the data set, only a limited number of strategies were considered and therefore a categorical variable was used.

9.2.3 Outcomes of interest

Four outcomes of interest were examined in this analysis. The outcomes of MI, stroke and all-cause mortality have been previously described (see 8.2.3). In the previous chapter, long term HbA1c was examined as a repeated measures outcome. In the present chapter, "achieving target HbA1c of 6.5%" was instead examined as a time to event outcome, as it was felt this would have a more useful clinical interpretation. For example, faced with a patient with an HbA1c just over 6.5%, it may be beneficial to have the flexibility to trial a lifestyle intervention in the knowledge that delaying treatment, perhaps to 8%, would not prolong the overall time to reaching their target. For this outcome, patients with initial HbA1c below or equal to 6.5% were excluded. As an initial descriptive analysis of this new outcome, crude incidence of achieving target HbA1c through time was calculated, and the associations between the covariates previously specified in 8.2.4 and outcome were examined via time to event analysis. As in the previous chapter (see sections 8.2.5.1 and 8.3.2) the association between covariates and outcome were considered in turn, and were adjusted for time-varying diabetes medication to ensure the association was not solely due to medication being on the causal pathway between the covariate and outcome. The range of strategies to be compared was restricted to 7, 8 9 and 10%, because for this particular outcome, the strategy of "treat when HbA1c raises

above 6.5%" would be equivalent to "treat from study entry" and therefore no longer dynamic.

9.2.4 Weighting models

9.2.4.1 Covariate selection

As with the standard MSM, application of dynamic MSMs relies on identifying a set of covariates that sufficiently adjust for confounding of the causal association between compliance to a particular strategy and outcome. As explained in section 4.5.3.3, the model for probability of compliance is equivalent to the model for probability of treatment initiation, but to avoid ambiguity, in this chapter this model will be referred to as the "weighting model" and the weights referred to as "inverse probability weights (IPWs)". For the outcomes of MI, stroke and all-cause mortality, which have already been examined in the previous chapter, the same set of covariates were used in the weighting models. Specifically, the models included age, gender, calendar period of diabetes onset (pre or post 2005), smoking status, alcohol consumption, use in the previous year of anti-hypertensive drugs, statins, NSAIDs or aspirin (baseline and time updated); previous history of any CVD, stroke, MI, CKD or cancer (baseline and time updated); HbA1c, BMI and SBP (baseline and time updated). These factors were also used in the previous chapter for the outcome of HbA1c trajectory (see sections 8.2.4 and 8.3.2), and are also likely to affect achieving target HbA1c of 6.5%. Based on this assumption and given the results of the initial descriptive analyses of observed associations with outcome (9.3.1), it was decided to use the same set of covariates for the analysis of time to target HbA1c.

9.2.4.2 Covariate form

Given the size of the data set and the need to use bootstrapping to obtain confidence intervals (as described in section 4.5.4.4), it was decided it would not be computationally practical to estimate weights or treatment effects for multiple covariate forms. In previous analyses (chapters 7 and 8), minimal differences in estimates of effect were observed between models using simple categorical variables, and more complex spline functions for continuous covariates. Therefore, for simplicity, only one set of covariate specifications was used in this analysis. In the weighting models, time since study entry and time between diagnosis and

study entry were entered as natural cubic splines with 5 knots (at 1, 10, 24, 48 and 120 months) and 4 knots (at 0, 1, 3 and 120 months) respectively, as previously specified (see 8.2.5.2.1 and appendix 20). HbA1c, BMI and SBP were categorised as in previous chapters, for both baseline and time updated values. Because age is strongly associated with cardiovascular outcomes in particular, age was categorised more finely than previously, in 10-year age bands between 45 and 75. <45 and >=75 remained as single categories due to a smaller number of patients in these age groups.

9.2.4.3 Weight estimation

Unstabilised weights were estimated in the non-expanded population as described in section 4.5.4. As the strategy of interest did not distinguish between metformin and sulfonylureas, the weighting model was fitted using logistic regression (outcome variable 0: no treatment, 1: initiated metformin or sulfonylureas). Due to negligible differences in estimates of effect when including inverse probability of censoring weights to account for informative censoring in previous chapters, censoring due to death, transfer out, end of data collection, or initiation of treatment other than metformin/sulfonylureas was not adjusted for using additional weighting in this analysis. This made the assumption that these censoring events were non-informative. Therefore, the resulting estimates could still be interpreted as the effects of each strategy that would have been observed had all patients initiated their first line therapy with metformin or a sulfonylurea in line with the strategy, and remained on one or both of these medications until the event of interest, or for 10 years, which was the maximum follow up period considered (see section 8.2.1). As in the previous chapter, the IPW's were truncated at the 99th percentile.

9.2.5 Addition of a grace period

As explained in section 4.5.5, greater compliance with the strategies of interest may be observed by extending the period in which treatment initiation can occur after a raised HbA1c measurement. In the analysis described above, a one month grace period was assumed. In the context of primary care, an additional grace period may be necessary to allow for factors such as patient indecision or delayed test results causing delays to treatment initiation.

HbA1c is a measure of blood glucose control over a time scale of approximately 3 months. Longer grace periods than this would increase the possibility that HbA1c (and other risk factors

for the outcome) could have changed by the end of the grace period. Therefore, 3 months was considered a reasonable choice for the maximum grace period. The decision of how far to extend the grace period was made by looking at how much compliance through time (the proportion of patients still at risk at a given time who were still compliant) was observed to increase for each strategy when extending the grace period in unit increments from 1 to 3 months. The additional grace period to be considered (either 2 or 3 months) was chosen as that which was observed to have the greatest increase in compliance compared to a one month grace period.

9.2.6 Fitting the dynamic MSM

For each treatment strategy identified (see section 9.3.2) each patient's data were replicated, and the patient censored at the end of the interval in which they became noncompliant with the strategy, as described in 4.5.4.3. Initially, compliance was defined with a one month grace period. The analysis was repeated allowing a grace period of three months, which was the value chosen based on the investigations described above (see section 9.3.3).

Since all outcomes were time to event outcomes, the dynamic Cox MSM was approximated using pooled logistic regression as outlined in section 4.5.4.4. The strategies for comparison were entered as a categorical variable, with the current guideline of 6.5% taken to be the reference category, except for the outcome of reaching target HbA1c, 7% was taken as the reference category since as previously explained, 6.5% was not included.

Time since study entry was modelled with a natural cubic spline. For MI, stroke and all-cause mortality, the same spline parameterisation from the previous chapter was used (specifically, a cubic spline with 5 knots at 1, 10, 24, 48 and 120 months). For time to target HbA1c, to allow more flexibility earlier in time (as suggested by initial investigations in 9.3.1), five evenly spaced percentile knots were used (1, 3, 6, 15, and 45 months).

The model included an interaction term between time since study entry and strategy. This was because, as explained in section 4.5.4.4, unless the risk of outcome is the same between the strategies, any difference between strategies cannot be constant through time. To provide easily interpretable estimates of the relative differences between strategies through follow up, a categorical term for time since study entry was used to model this interaction. This parameterisation for time since study entry was included the model in addition to the cubic spline parameterisation described above. The relative effect of each strategy on risk of each

outcome was estimated for 0-6 months, >6-12months, >1-2 years, >2-3 years, >3-4 years and >4 years after study entry. These categorisations were initially made with the outcome of time to target HbA1c in mind, as this is a shorter term outcome, and the number of events observed after 4 years of follow up was low (see 9.3.1). Splitting the first year into two 6 month intervals was also relevant for the CV outcomes to detect possible residual confounding due to immediate risk of the event close to the time of diabetes diagnosis. For example, this could occur if immediate risk of death not captured by the covariates in the model had prevented a patient from initiating treatment regardless of their HbA1c.

As well as estimation of relative effects of strategies, cumulative incidence at 1, 2 and 4 years post study entry was estimated for each strategy. To calculate smooth cumulative incidence curves to 10 years, a separate model was fitted where the categorical form of time was removed, and the interaction between time and strategy was fitted using the cubic spline term for time.

For each outcome, the dynamic MSM included only time, treatment strategy and the interaction between them (parameterised as described above), since the unstabilised weight also adjusts for baseline covariates. To assess the impact of the weighting on the estimated effects of the dynamic strategies, an unweighted model was also fitted in the expanded data, with adjustment for baseline covariates. This will be referred to as the "unweighted analysis".

Confidence intervals for the hazard ratios and cumulative incidence at 1, 2 and 4 years from study entry were obtained via bootstrapping with 200 replications (see 4.5.4.4), stratified by calendar year of diagnosis. The cumulative incidence curves were plotted without confidence intervals. To reduce computation time, confidence intervals were not obtained for the cumulative incidence estimates in the unweighted analyses.

9.2.7 Sensitivity analyses

The sensitivity analyses described below were carried out for the dynamic MSM only with a one-month grace period.

9.2.7.1 Defining compliance as initiation of metformin only

The overall strategy of interest for the main analysis did not distinguish between initiating with metformin or a sulfonylurea. To check whether possible differing effects of metformin and

sulfonylureas on all outcomes may alter the estimated effects of the strategies, the first sensitivity analysis re-defined the strategies as "treat with metformin once HbA1c rises above x%", for the same values of x as in the main analysis. For simplicity, initiation of a sulfonylurea was treated as a separate censoring event to compliance. It was coded in the same way as the other censoring events of death (if not the outcome), transfer from practice, and initiation of diabetes medications other than metformin or sulfonylureas, in that if it occurred, the patient was considered as if lost to follow up from the beginning of that interval. The additional censoring for sulfonylurea use was done in two ways for each outcome. Firstly, an **ITT style approach** was taken, whereby, as per the main analysis in the previous chapter (see section 8.2.2), if a patient initiated metformin and then subsequently initiated a sulfonylurea, they were assumed to be still exposed to metformin and remained in the study until they were no longer considered to be at risk (as defined in 9.2.1). Therefore, the patient was only censored if initiating sulfonylureas as a first line therapy. This broadly redefined the strategy to be "initiate with metformin once HbA1c rises above x%, then continue with metformin, sulfonylureas or both as necessary". Secondly, an **as treated style approach** was taken whereby patients were censored at any initiation of a sulfonylurea during follow up (as in section 8.2.5.3.5), effectively redefined the strategy as "treat with metformin monotherapy when HbA1c first rises above x% and maintain monotherapy". For both approaches, the IPW were re-estimated by modelling the probability of treatment (none, metformin) as described in sections 4.5.3.3 and 4.5.4.2. Since it was felt that in both approaches, sulfonylurea initiation was unlikely to be non-informative, separate IPCW for censoring due to sulfonylurea initiation were estimated as outlined in section 4.5.3.4. For the ITT style approach, since the probability of censoring due to sulfonylurea initiation becomes zero once metformin is initiated, the IPCW model was fitted only in patients who were still untreated in the previous interval, and the probability of remaining uncensored was set to 1 from the interval after which metformin was initiated.

9.2.7.2 Using a HbA1c target of 6% instead of 6.5%

Using 6.5% as the target HbA1c meant that the treatment threshold of 6.5% was included in the main analysis. In order to look at this threshold in comparison to higher values of HbA1c with regard to speed of glucose lowering, the outcome was re-defined as obtaining an HbA1c of 6%.

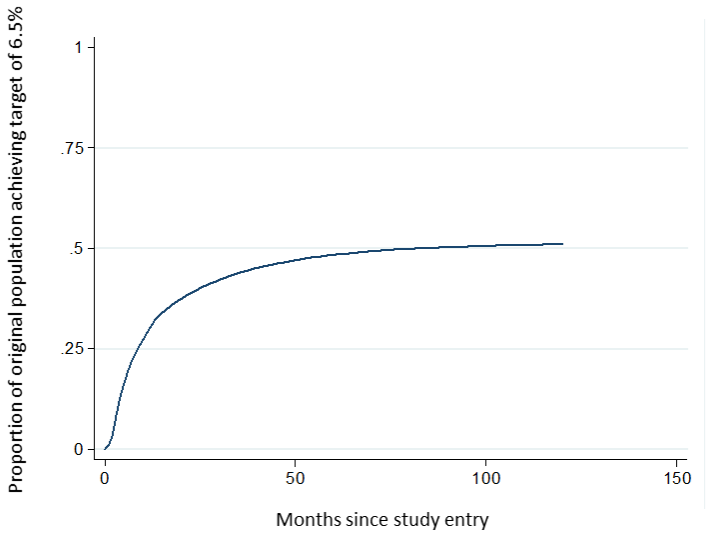
9.3 RESULTS

9.3.1 Descriptive analysis of outcome: Achieving target HbA1c of 6.5%

Removing patients from the cohort who had an HbA1c less than or equal to 6.5% at study entry left 31,872 patients, of which 131 were censored due to death, transfer out of practice or initiation of medication other than metformin or sulfonylureas in month one, and therefore did not contribute to the analysis. This left 31,741 patients. The general demographics for this reduced population, aside from the differing distribution of baseline HbA1c, were similar to that presented in the previous chapter (see section 8.3.1 and Table 8.1) and are presented in appendix 25 table 25.1. The crude rate of achieving target HbA1c was 18.2 per 1000 person months, 95% CI (17.8 – 18.5). Figure 9.1 shows the observed cumulative incidence (as a proportion of the initial patient population) for achieving target through time.

Figure 9.2 shows the crude probability of target attainment per month of follow up in those still at risk, with 95% confidence bars, for the whole of follow up (top) and for the first 50 months (bottom).

Figure 9.1 Proportion of original population achieving target HbA1c of 6.5% through follow up



Month	0	20	40	60	80	100	120
N still at risk	31741	14142	8057	4715	2728	1437	708

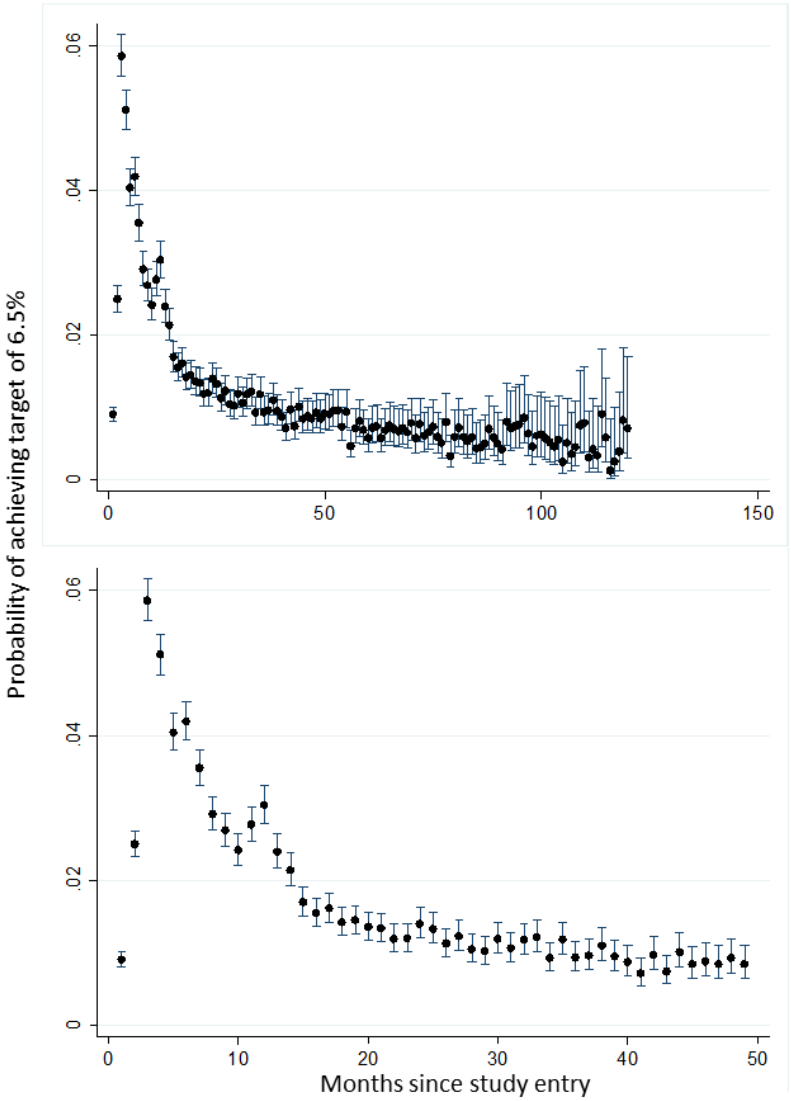
In the overall population, the highest probabilities of reaching target were in the first 12 months after study entry, with this declining over time. After about 50 months, the probabilities were relatively constant with confidence intervals widening as less patients were still at risk, and fewer events occurred.

Observed associations (estimated HR and 95% CI adjusted for time updated diabetes medication use) between all covariates (where each covariate was considered in turn) and target HbA1c are displayed in Table 9.1. Older age was associated with an increased likelihood of reaching target HbA1c. Patients diagnosed with diabetes after 2005 were slightly less likely to reach target compared to those diagnosed between 2000 and 2005. Current smokers were less likely to reach target compared to non-smokers, but heavier drinkers tended to be more likely to reach target. As would be expected, the higher the baseline/previous HbA1c, the less likely the patient was to achieve their target. Time updated BMI and SBP appeared to have a similar direction of association to HbA1c for both baseline and time updated values, but the size of the effect was smaller.

Use of antihypertensive drugs or aspirin, both at baseline and in the previous year, was associated with increased likelihood of reaching target, as were time updated recent history of MI, stroke, and other CVD events, and ever history of CKD and cancer. On the other hand, recent use of statins was associated with a decreased chance of reaching target Hba1c

For some baseline indicators of underlying health such as use of statins and use of NSAIDs, and ever history of CVD events, there was no strong evidence of an association with reaching target HBA1c. However, they were still included in the models for the weights, as was gender. This was because as with previous chapters, a conservative approach which reduces the chance of omitting important risk factors in error was preferred. Additionally, these covariates have not previously been shown to have strong associations with treatment, and so the risk of violations of the positivity assumption from their inclusion was considered minimal.

Figure 9.2 Probability of achieving target HbA1c in each month interval from study entry, for patients still under follow up and yet to achieve target.



Top figure shows full 10 years of follow up, bottom figure shows first 50 months only for clarity.

Risk Factor	Hazard Ratio	95% CI
Age		
<45	1 (ref)	
45-54	1.06	(1.00 , 1.13)
55-64	1.24	(1.17 , 1.31)
65-74	1.42	(1.34 , 1.50)
75+	1.43	(1.35 , 1.52)
Gender		
Male	1 (ref)	
Female	1.01	(0.98 , 1.03)
Smoking status		
Non	1 (ref)	
Current	0.86	(0.83 , 0.90)
Ex	1.05	(1.02 , 1.08)
Drinking Status		
Non drinker	1 (ref)	
ex-drinker	1.11	(1.04 , 1.18)
current drinker unknown	1.02	(0.91 , 1.14)
rare drinker <2u/d	1.05	(1.00 , 1.10)
moderate drinker 3-6u/d	1.13	(1.08 , 1.18)
excessive drinker >6u/d	1.19	(1.12 , 1.27)
Year of diabetes onset		
2000-2005	1 (ref)	
post 2005	0.90	(0.87 , 0.93)
Baseline BMI		
<25	1 (ref)	
25-29	1.07	(1.02 , 1.13)
30-34	1.05	(1.00 , 1.11)
35+	1.00	(0.95 , 1.06)
Baseline HbA1c		
6.5 -7%	1 (ref)	
7-8%	0.55	(0.53 , 0.57)
8-10%	0.41	(0.39 , 0.43)
10% +	0.34	(0.32 , 0.35)
Baseline SBP		
100-129	1 (ref)	
130-139	1.04	(1.00 , 1.07)
140-149	1.09	(1.02 , 1.16)
150+	1.05	(1.01 , 1.09)
Use in year before baseline of…		
Anti-hypertensive drugs	1.21	(1.17 , 1.24)
Statins	1.01	(0.98 , 1.04)
NSAIDS	1.03	(0.99 , 1.07)
Aspirin	1.06	(1.02 , 1.09)
Event in three months before baseline of…		
Any CVD	1.08	(0.97 , 1.20)
MI	0.74	(0.56 , 1.00)
Stroke	1.13	(0.9 , 1.41)
Any history at baseline of …		
CVD	1.01	(0.97 , 1.06)
MI	0.93	(0.86 , 1.01)
Stoke	1.04	(0.96 , 1.12)
CKD	1.22	(1.15 , 1.30)
Cancer	0.82	(0.56 , 1.20)

Table 9.1 Associations between all covariates (considered in turn) and reaching target HBA1c.

Estimated from analyses adjusted for time-dependent use of metformin and sulfonylureas. Hazard ratio represents relative risk of achieving target of 6.5% compared to reference group. For variables without a reference indicated, reference is no history/no use.

Risk Factor	Estimate	95% CI
Previous BMI (interval -1)		
<25	1 (ref)	
25-29	0.97	(0.93 , 1.02)
30-34	0.89	(0.85 , 0.93)
35+	0.82	(0.78 , 0.87)
Previous HbA1c (interval -1)		
6.5 -7%	1 (ref)	
7-8%	0.44	(0.42 , 0.45)
8-10%	0.29	(0.27 , 0.30)
10% +	0.21	(0.2 , 0.23)
Previous HbA1c (interval -2)		
6.5 -7%	1 (ref)	
7-8%	0.43	(0.42 , 0.44)
8-10%	0.27	(0.26 , 0.29)
10% +	0.19	(0.18 , 0.21)
Previous SBP (interval -1)		
100-129	1 (ref)	
130-139	0.94	(0.90 , 0.97)
140-149	0.92	(0.88 , 0.95)
150+	0.82	(0.78 , 0.86)
Use in previous year of..		
Anti-hypertensive drugs	1.19	(1.16 , 1.23)
Statins	0.95	(0.93 , 0.98)
NSAIDS	1.06	(1.02 , 1.10)
Aspirin	1.08	(1.05 , 1.12)
Event in previous three months of…		
Any CVD	1.36	(1.19 , 1.56)
MI	1.43	(1.03 , 1.98)
Stroke	1.70	(1.31 , 2.22)
Any history of …		
CVD	1.06	(1.02 , 1.11)
MI	1.00	(0.92 , 1.07)
Stoke	1.10	(1.02 , 1.19)
CKD	1.28	(1.21 , 1.34)
Cancer	1.34	(1.20 , 1.50)

Table 9.1 continued: Associations between all covariates (considered in turn) and reaching target HBA1c.

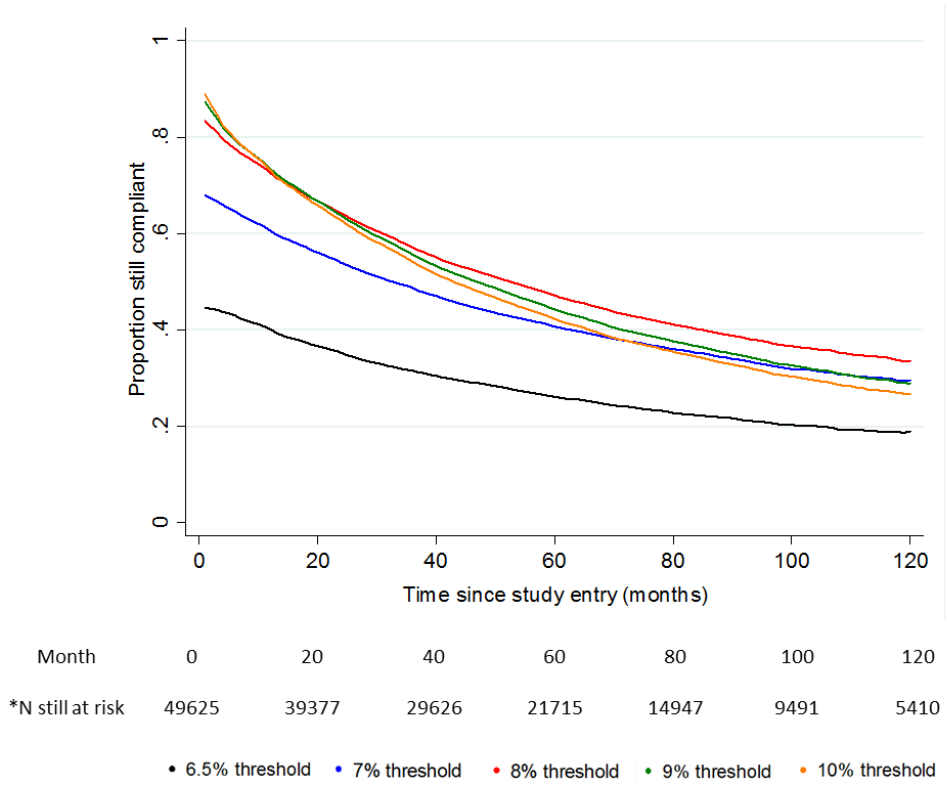
Estimated from analyses adjusted for time-dependent use of metformin and sulfonylureas. Hazard ratio represents relative risk of achieving target of 6.5% compared to reference group. For variables without a reference indicated, reference is no history/no use

.

9.3.2 Defining the set of plausible strategies

Compliance to different treatment strategies was examined as described in section 9.2.2 in the study population outlined in section 9.2.1 (that included patients with baseline Hba1c less than or equal to 6.5%, where patients were followed up to death, transfer out, end of data collection or initiation of any treatment other than metformin or sulfonylureas). This was to examine compliance in the overall population, and not just that relevant to the outcome of achieving target Hba1c. The numbers remaining compliant as a proportion of the patients still considered at risk through time (i.e. not censored due to death, end of data collection, transfer out or initiation of other medication) are shown in Figure 9.3 for thresholds of 6.5, 7, 8, 9 and 10%. Similar results were found when excluding follow up after the first HbA1c of 6.5% or less (the endpoint for the time-to -target analysis), and are presented in appendix 25, figure 25.1. Table 9.2 shows that all strategies had sufficient overall compliance to be compared and had variation in reasons for (non)compliance.

Figure 9.3 Proportion of at risk patients who were still compliant to the treatment strategy "treat with metformin or sulfonylureas when HbA1c first rises above x%" at each month of follow up, for x = 6.5, 7, 8, 9 and 10.



*N at risk at time 0 excludes 125 patients censored for transfer out of practice or initiation of medication other than metformin or sulfonylureas in month1.

	Compliant			Non - compliant		
Treatment threshold	**Remain untreated**	**Treated over threshold**	**Total % compliant**	**Treated below threshold**	**Not treated above threshold**	**Total % on-compliant**
6.5%	15.5%	10.8%	**26.2%**	2.8%	70.9%	**73.8%**
7%	32.0%	12.9%	**44.9%**	9.5%	45.7%	**55.1%**
8%	41.9%	10.9%	**52.8%**	29.4%	17.9%	**47.2%**
9%	43.8%	6.9%	**50.7%**	40.2%	9.1%	**49.3%**
10%	44.5%	4.4%	**49.0%**	45.9%	5.2%	**51.0%**

Table 9.2 Percentage of patients who remain, or do not remain compliant with each strategy for their time at risk up to and including first initiation with metformin or a sulfonylurea.

Example: A patient initiates a sulfonylurea in the interval after their HbA1c raises to 8.5% from 6.8%. This patient contributes to "noncompliant, not treated above threshold" for the strategy of 6.5%, "compliant, treated over threshold" for 7% and 8%, and "noncompliant, treated below threshold" for 9 and 10%. If the same patient initiated insulin instead, they would be censored from the beginning of that interval, and so no longer at risk to become compliant by being treated for 7% and 8%, or noncompliant with the 9% and 10% strategies. Instead, they would be considered "compliant, remain untreated" for all of the 7, 8, 9 and 10% strategies, but still "noncompliant not treated above threshold" for the 6.5% strategy.

9.3.3 Addition of a grace period for time allowed between HbA1c exceeding threshold and treatment initiation

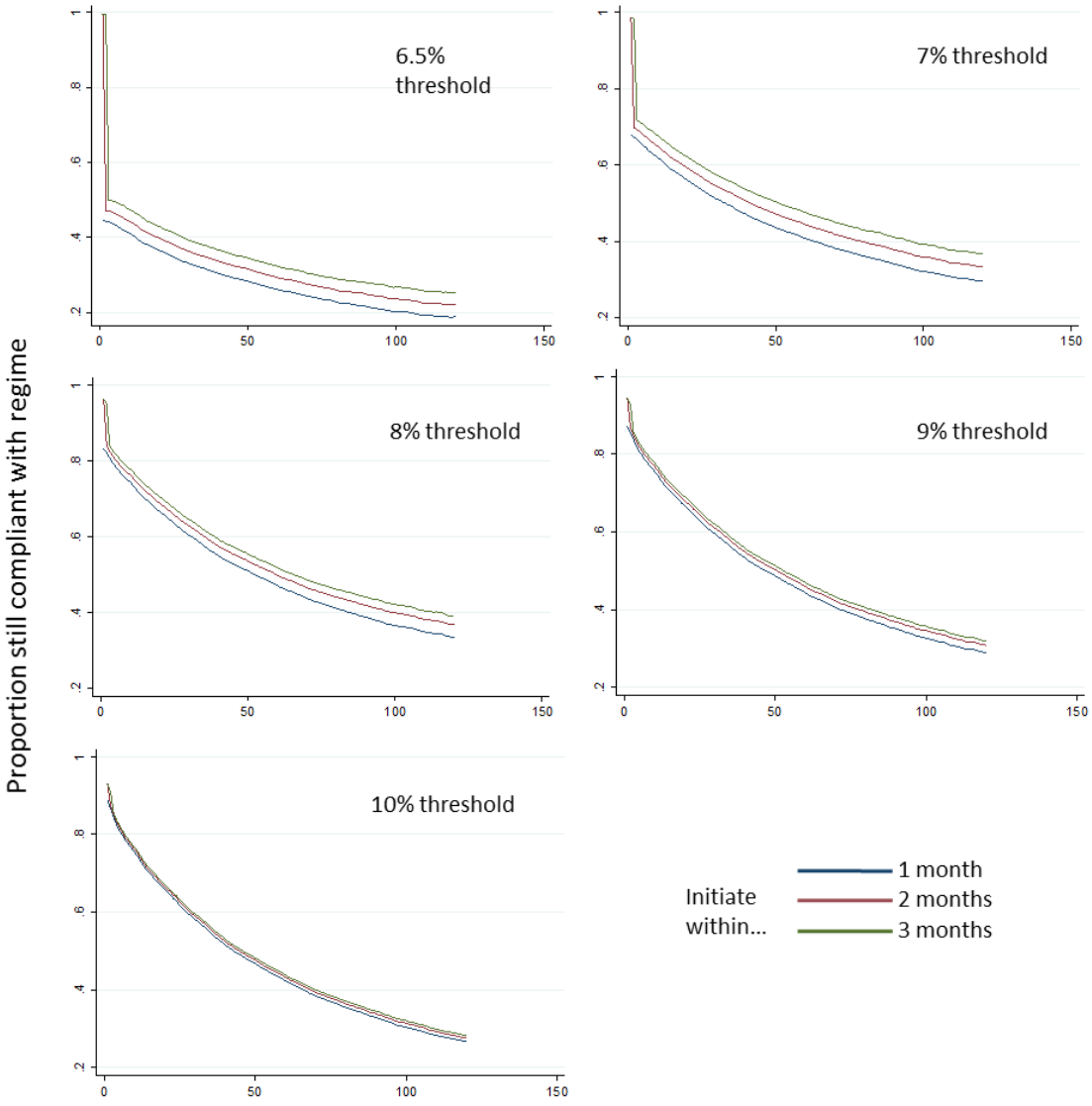
Compliance according to varying grace period used is presented for each treatment strategy in Figure 9.4 as the proportion of patients at risk who were complaint with the strategy at each month of follow up. For the stricter strategies (lower HbA1c thresholds), compliance was improved almost uniformly for every extra month of grace period, but only by a small amount. As the HbA1c threshold increased, there was even less compliance gained by allowing for a longer grace period. This is likely to be because for these strategies, non compliance due to initiating treatment before the threshold is common, and this is not affected by adding the grace period. However, since the greatest increase in compliance was observed for a grace period of three months, this was the length of grace period chosen for the secondary analysis, with a one month period for initiation kept for the primary analysis.

1.1.1 Calculating IPW

For each outcome, and for both one and three month grace periods, the distributions of IPWs before and after truncation at the 99th percentile are presented in Table 9.3. As expected, the distributions were very similar across outcomes, but are shown for each outcome for clarity due

to differences in follow up time. Due to large maximum weights before truncation, the truncated weights were used to weight the dynamic MSMs.

Figure 9.4 Proportion of patients at risk still compliant to the treatment strategy for different lengths of grace period.



Month	0	20	40	60	80	100	120
N still at risk	49625	39377	29626	21715	14947	9491	5410

*N at risk at time 0 excludes 125 patients censored for transfer out of practice or initiation of medication other than metformin or sulfonylureas in month1.

	Target HbA1c of 6.5%				MI			
	One month grace period		Three month grace period		One month grace period		Three month grace period	
	Untruncated	Truncated	Untruncated	Truncated	Untruncated	Truncated	Untruncated	Truncated
Mean	4.61	2.9	3.3	2.1	2.8	2.1	2.2	1.6
SD	115.3	3.7	103.4	2.8	41.6	2.7	34.9	2.2
25th Percentile	1.1	1.1	1.0	1.0	1.1	1.1	1.0	1.0
50th Percentile	1.5	1.5	1.1	1.1	1.1	1.1	1.1	1.1
75th Percentile	3.3	3.3	1.6	1.6	1.7	1.7	1.3	1.3
Max	16209.4	27.0*	16209.4	20.0*	17011.2	18.9*	17011.2	17.1
	Stroke				**All- Cause mortality**			
	One month grace period		**Three month grace period**		**One month grace period**		**Three month grace period**	
	Untruncated	Truncated	Untruncated	Truncated	Untruncated	truncated	Untruncated	Truncated
Mean	2.8	2.1	2.1	1.6	2.8	2.1	2.1	1.7
SD	40.7	2.7	32.7	2.2	41.3	2.8	33.9	2.2
25th Percentile	1.1	1.1	1.0	1.0	1.1	1.1	1.0	1.0
50th Percentile	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
75th Percentile	1.7	1.7	1.3	1.3	1.7	1.7	1.3	1.3
Max	17692.3	18.8*	17692.3	17.1*	18032.5	19.0*	18032.5	17.1*

Table 9.3 Distribution of untruncated and truncated unstabilised IPWs for each outcome model, for strategies allowing both a one month and three month grace period for treatment initiation.

*99th percentile of untruncated weight distribution

9.3.4 Comparison of dynamic strategies

9.3.4.1 Target HbA1c of 6.5%

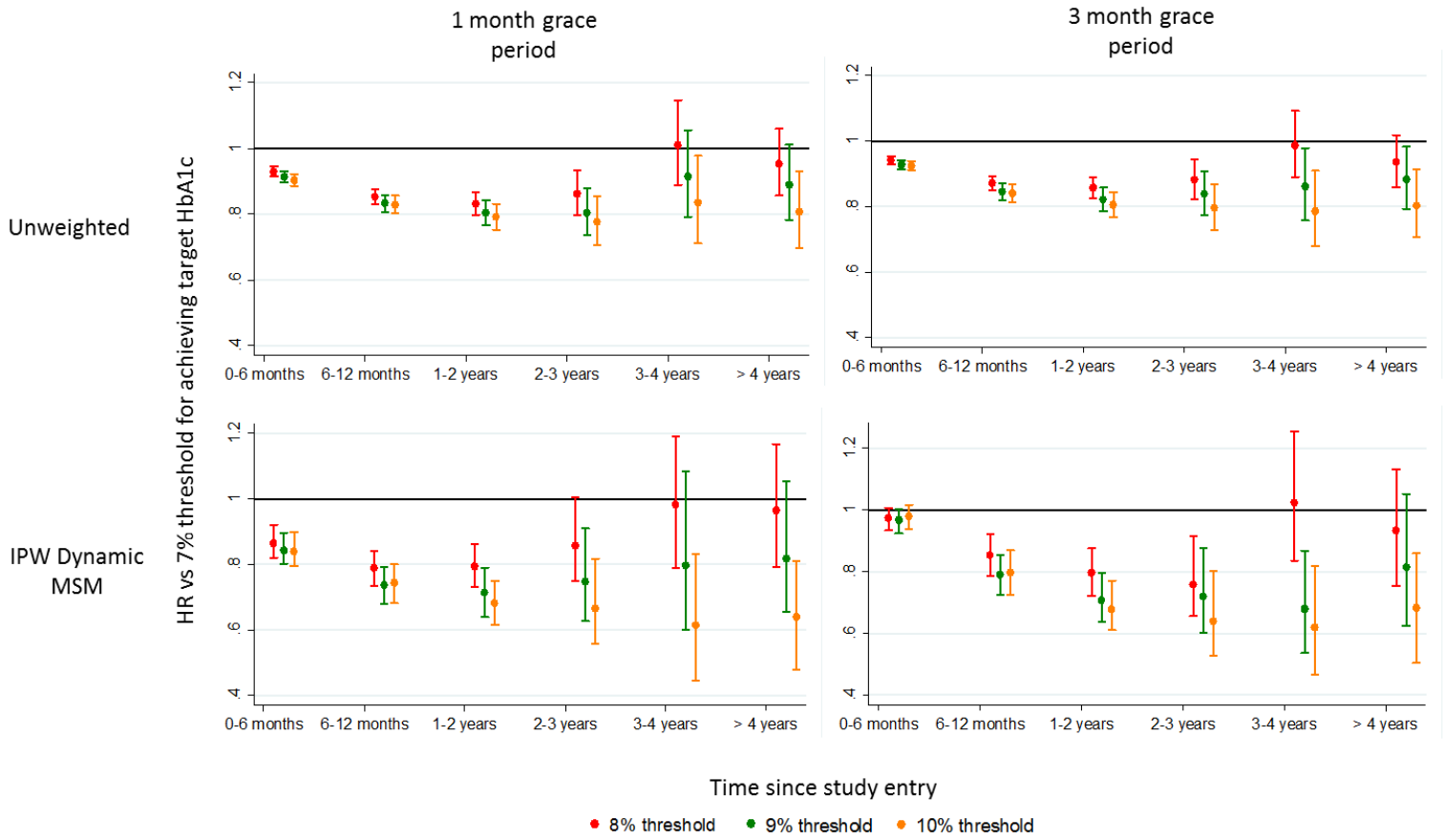
In both unweighted and weighted models, higher thresholds of HbA1c were associated with lower rates of reaching target HbA1c of 6.5% compared to the 7% threshold. In the dynamic MSM; the reduction in target attainment vs the 7% threshold was generally of greater magnitude than in the unweighted analysis, particularly for the strategies of 9% and 10%. Table 9.4 displays the estimated hazard ratios for each strategy vs a 7% initiation threshold, for the unweighted model (top) and the dynamic MSM (bottom), for models assuming a one-month grace period. The same estimates and 95% confidence intervals are displayed graphically in Figure 9.5. For all time periods, the precision of the estimates was lower for the dynamic MSM than for the unweighted analysis. This was most prominent for periods > 2 years from study entry.

	Unweighted model baseline adjusted – hazard ratio for strategy vs 7% for reaching target HbA1c: one month grace period (HR<1 indicates inferior strategy)					
Strategy threshold	**0-6 months[1]**	**6-12 months[1]**	**1-2 years[1]**	**2-3 years[1]**	**2-4 years[1]**	**>4 years[1]**
7%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
8%	0.93 (0.92 , 0.95)	0.85 (0.83 , 0.88)	0.83 (0.80 , 0.86)	0.86 (0.81 , 0.93)	1.01 (0.90 , 1.17)	0.95 (0.87 , 1.05)
9%	0.91 (0.90 , 0.93)	0.83 (0.81 , 0.86)	0.80 (0.77 , 0.84)	0.80 (0.75 , 0.87)	0.91 (0.80 , 1.08)	0.89 (0.80 , 1.00)
10%	0.90 (0.89 , 0.92)	0.83 (0.80 , 0.85)	0.79 (0.76 , 0.83)	0.78 (0.71 , 0.85)	0.83 (0.71 , 1.02)	0.81 (0.70 , 0.91)
	IPW* Dynamic MSM – hazard ratio for strategy vs 7% for reaching target HbA1c: one month grace period (HR<1 indicates inferior strategy)					
Strategy threshold	**0-6 months[1]**	**6-12 months[1]**	**1-2 years[1]**	**2-3 years[1]**	**2-4 years[1]**	**>4 years[1]**
7%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
8%	0.86 (0.82 , 0.92)	0.79 (0.73 , 0.84)	0.79 (0.73 , 0.86)	0.86 (0.75 , 1.00)	0.98 (0.79 , 1.19)	0.96 (0.79 , 1.17)
9%	0.84 (0.80 , 0.90)	0.74 (0.68 , 0.79)	0.71 (0.64 , 0.79)	0.75 (0.63 , 0.91)	0.8 (0.60 , 1.08)	0.82 (0.66 , 1.05)
10%	0.84 (0.80 , 0.90)	0.74 (0.68 , 0.80)	0.68 (0.61 , 0.75)	0.67 (0.56 , 0.82)	0.61 (0.45 , 0.83)	0.64 (0.48 , 0.81)

Table 9.4 Hazard ratios (and 95% CIs) to compare strategy of "treat in the interval following that when HbA1c exceeds x%" for X = 8, 9, 10 and reference strategy of x=7 in terms of reaching target HbA1c of 6.5%.

[1]Columns show estimated HR and 95% CI by time since study entry. CI's obtained via 200 bootstrap replications
* weighting model includes: age gender, calendar period of diabetes onset (pre or post 2005), smoking status, alcohol consumption, use in the previous year of anti-hypertensive drugs, statins, NSAIDS or aspirin (baseline and time updated); previous history of any CVD, stroke, MI, CKD or cancer (baseline and time updated); HbA1c, BMI and SBP (baseline and time updated).

Figure 9.5 Estimated HRs and 95% confidence intervals to compare target HbA1c attainment through time for different HbA1c thresholds for treatment initiation vs a 7% threshold.



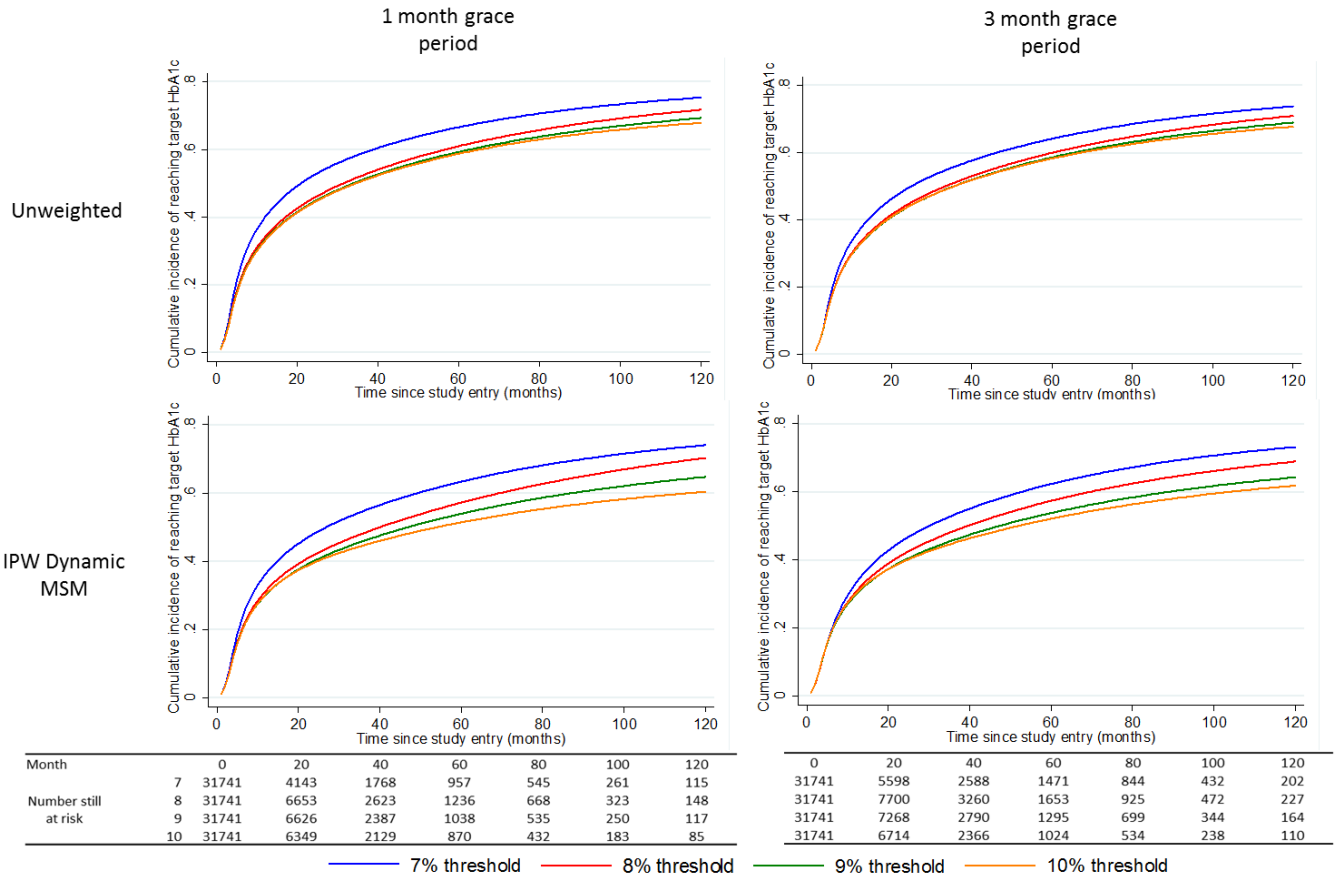
Reference category = 7% threshold for initiation. Results from unweighted model (top) and Dynamic MSM (bottom) allowing one month grace period (left) and three month grace period (right) for treatment initiation. CI's obtained via 200 bootstrap replications

Allowing a grace period of three months instead of one month made no major differences to either the unweighted model or the dynamic MSM, with the exception of 0-6 months from study entry, where for the dynamic MSM the estimates were slightly closer to those from the unweighted model compared to when using one month (Figure 9.5, and appendix 26, table 26.1).

Figure 9.6 shows the estimated cumulative incidence curves for target HbA1c attainment for each of the 4 strategies of interest. The dose response effect estimated by the dynamic MSM seen in Table 9.4 (and in table 26.1 in appendix 26) is reflected by the cumulative incidence curve, suggesting that higher thresholds for treatment initiation are associated with lower rates of reaching target HbA1c. In the dynamic MSM, by 4 years, under the strictest strategy (treat at 7%) 59% (95% CI 58%-61%) of the population were estimated to reach target, compared with 48% at 4 years (95% CI 47% -50%) under the most lenient strategy (treat at 10%). The estimated cumulative incidences from the unweighted analyses were estimated to be higher than those from the MSM for all time points and for all strategies (Table 9.5).

To further understand the finding of decreased incidence of achieving target HbA1c with the higher thresholds, the percentage of weighted person time in particular ranges of HbA1c up to the interval before target HbA1c was reached were estimated separately for each strategy and separately for time on and off treatment. These proportions are presented in Table 9.6. Once treated, the proportion of time at low range of HbA1c was highest for the stricter treatment strategies. For a threshold of 7%, 27% of treated person time was in the HBA1c range 6.5- 7%. The equivalent proportion for a threshold of 10 was 15%. In addition, for the threshold of 10%, another 25% of the treated person time was still spent with HbA1c >10%. This suggests that the effect of treatment is not immediate and that the lower rate of target attainment for the higher thresholds is not solely due to waiting longer to initiate treatment while HbA1c rises.

Figure 9.6 Cumulative incidence curves for reaching target HbA1c of 6.5% or less, for different HbA1c thresholds for treatment initiation.



Curves are estimated from unweighted models adjusting for baseline covariates (top), and dynamic MSM with IPW (bottom). Curves are estimated or a allowing a one month grace period for initiation (left) three month grace period (right)

		Unweighted (one-month grace period): proportion achieving target HBA1c by…			IPW Dynamic MSM (one-month grace period): proportion achieving target HBA1c by…		
		1 year	**2 years**	**4 years**	**1 year**	**2 years**	**4 years**
HbA1c threshold	7%	0.40	0.53	0.63	0.36 (0.35 , 0.38)	0.49 (0.47 , 0.50)	0.59 (0.58 , 0.61)
	8%	0.34	0.46	0.57	0.32 (0.31 , 0.32)	0.42 (0.41 , 0.43)	0.53 (0.52 , 0.54)
	9%	0.33	0.45	0.56	0.30 (0.30 , 0.31)	0.40 (0.39 , 0.41)	0.50 (0.49 , 0.52)
	10%	0.33	0.45	0.55	0.30 (0.30 , 0.31)	0.40 (0.39 , 0.41)	0.48 (0.47 , 0.50)

Table 9.5 Estimated proportions of population achieving target HbA1c by 1, 2 and 4 years from study entry, for each treatment strategy.

95% CI given in brackets for dynamic MSM only, obtained via 200 bootstrap replications.

	Proportion of pre and post treatment weighted person time (using truncated weights) within given HbA1c range for treatment strategies of…							
	7%		**8%**		**9%**		**10%**	
HbA1c range (%)	Pre-treat	Post -treat	Pre-treat	Post - treat	Pre-treat	Post -treat	Pre-treat	Post - treat
6.5-7	83.7%	26.5%	49.6%	19.5%	40.6%	16.3%	37.5%	15.1%
7-7.5	7.1%	28.1%	31.2%	22.6%	27.6%	19.1%	25.7%	16.3%
7.5-8	3.0%	17.0%	14.4%	15.9%	14.6%	15.3%	14.1%	13.4%
8-8.5	1.6%	8.4%	1.7%	12.3%	8.8%	10.1%	8.9%	9.5%
8.5-9	1.0%	5.5%	0.9%	8.4%	5.7%	8.1%	5.8%	8.4%
9-9.5	0.8%	3.8%	0.5%	5.6%	0.9%	8.5%	3.8%	7.0%
9.5-10	0.6%	2.8%	0.4%	4.2%	0.6%	6.1%	2.6%	4.7%
10-11	1.0%	3.8%	0.5%	5.6%	0.6%	7.9%	0.8%	12.3%
11+	1.3%	4.0%	0.7%	5.8%	0.7%	8.7%	0.8%	13.2%
total	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Total weighted person months								
Truncated	178,086	1,087,782	496,788	765,828	635,135	608,913	712,374	394,619
Un-truncated	170,432	1,141,529	909,690	876,218	1,136,210	1,241,113	1,411,086	681,093

Table 9.6 Proportion of pre and post treatment person time at HbA1c range for treatment strategies of 7%, 8%, 9% and 10%, estimated from the weighted population.

Greyed out numbers indicate ranges where HbA1c is above the threshold for that strategy during time off treatment. Each patient can contribute a maximum of two person months towards these percentages (i.e. the month before treatment initiation and the month in which treatment is (not) initiated).

9.3.4.2 MI

For the first 6 months of follow up, the dynamic MSM with a one-month grace period estimated that all treatment strategies reduced the risk of MI compared to a 6.5% threshold, albeit with wide confidence intervals. From 6 months onwards, there was a suggestion of increasing risk of MI with the higher HbA1c thresholds but again the estimates had low precision (Table 9.7, bottom, Figure 9.7 bottom left). In the unweighted analysis, the increased risk from 6 months to 2 years after study entry was not apparent (Table 9.7, top, Figure 9.7 top left).

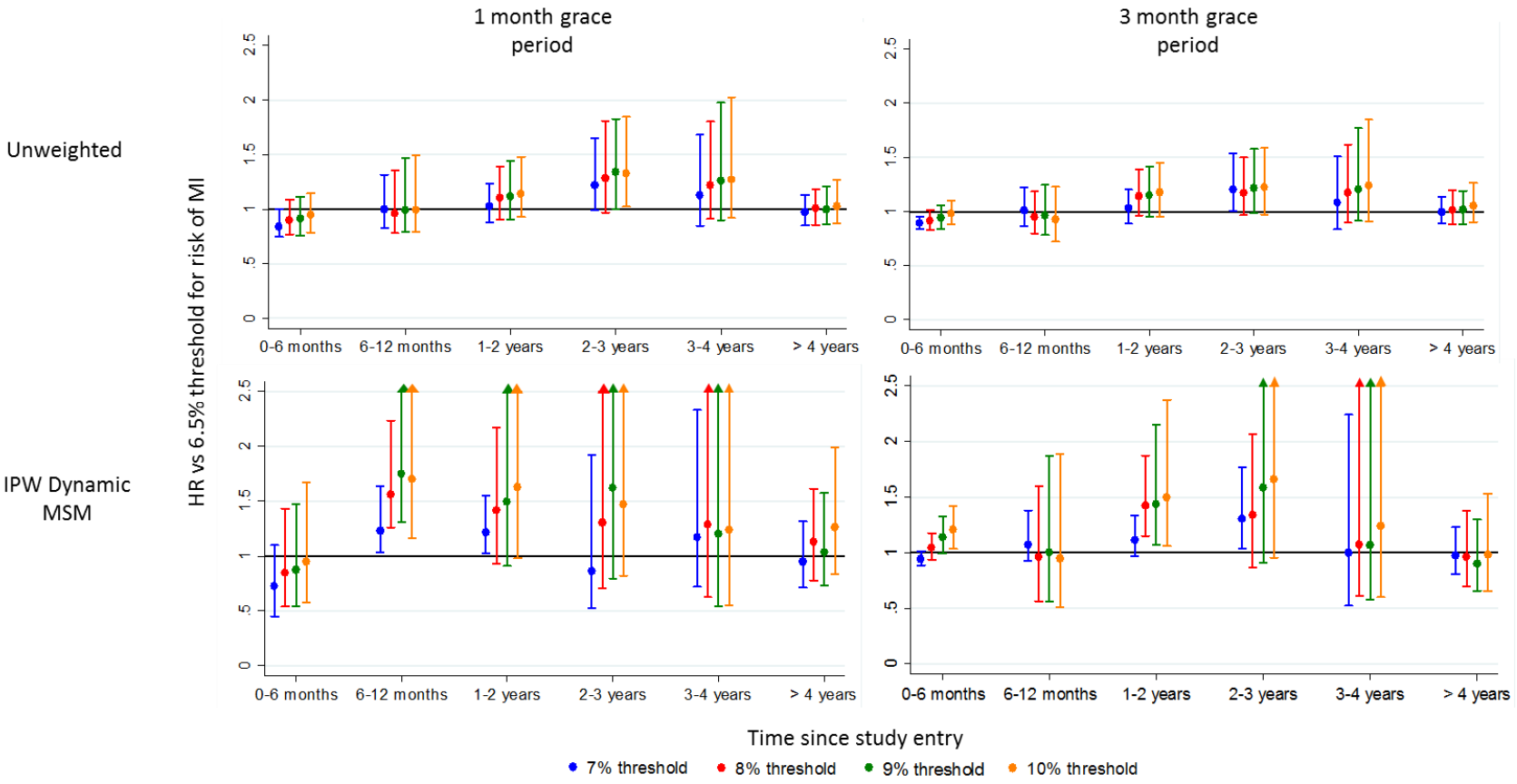
	Unweighted model baseline adjusted – hazard ratio for strategy vs 6.5% for risk MI: one month grace period (HR<1 indicates superior strategy)					
Strategy threshold	**0-6 months[1]**	**6-12 months[1]**	**1-2 years[1]**	**2-3 years[1]**	**2-4 years[1]**	**>4 years[1]**
6.5%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
7%	0.84 (0.75 , 1.00)	1.00 (0.83 , 1.31)	1.03 (0.88 , 1.24)	1.22 (0.99 , 1.65)	1.13 (0.85 , 1.68)	0.97 (0.85 , 1.13)
8%	0.90 (0.77 , 1.09)	0.96 (0.78 , 1.35)	1.10 (0.91 , 1.39)	1.29 (0.97 , 1.81)	1.22 (0.91 , 1.80)	1.01 (0.86 , 1.18)
9%	0.91 (0.76 , 1.11)	0.99 (0.79 , 1.47)	1.12 (0.91 , 1.44)	1.34 (1.00 , 1.82)	1.26 (0.90 , 1.97)	1.00 (0.86 , 1.21)
10%	0.95 (0.79 , 1.14)	0.99 (0.80 , 1.50)	1.14 (0.93 , 1.48)	1.33 (1.02 , 1.84)	1.27 (0.92 , 2.02)	1.03 (0.86 , 1.27)
	IPW* Dynamic MSM – hazard ratio for strategy vs 6.5% for risk MI: one month grace period (HR<1 indicates superior strategy)					
Strategy threshold	**0-6 months[1]**	**6-12 months[1]**	**1-2 years[1]**	**2-3 years[1]**	**2-4 years[1]**	**>4 years[1]**
6.5%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
7%	0.73 (0.45 , 1.10)	1.23 (1.03 , 1.63)	1.22 (1.02 , 1.55)	0.86 (0.53 , 1.92)	1.17 (0.72 , 2.33)	0.95 (0.71 , 1.32)
8%	0.84 (0.54 , 1.43)	1.56 (1.26 , 2.24)	1.42 (0.93 , 2.16)	1.30 (0.70 , 3.28)	1.29 (0.62 , 3.66)	1.13 (0.78 , 1.61)
9%	0.87 (0.55 , 1.48)	1.74 (1.31 , 2.69)	1.49 (0.91 , 2.49)	1.62 (0.79 , 3.82)	1.20 (0.54 , 3.31)	1.03 (0.73 , 1.57)
10%	0.95 (0.58 , 1.67)	1.70 (1.16 , 2.89)	1.63 (0.98 , 2.81)	1.47 (0.82 , 4.14)	1.24 (0.55 , 3.86)	1.26 (0.84 , 1.99)

Table 9.7 Hazard ratios (and 95% CIs) to compare risk of MI between strategy of "treat in the interval following that when HbA1c exceeds x%" for X = 7, 8, 9, 10 and reference strategy of x=6.5.

[1]`Columns show estimated HR and 95% CI by time since study entry. CI's obtained via 200 bootstrap replications.
* weighting model includes: age gender, calendar period of diabetes onset (pre or post 2005), smoking status, alcohol consumption, use in the previous year of anti-hypertensive drugs, statins, NSAIDS or aspirin (baseline and time updated); previous history of any CVD, stroke, MI, CKD or cancer (baseline and time updated); HbA1c, BMI and SBP (baseline and time updated).

.

Figure 9.7 Estimated HRs and 95% confidence intervals to compare risk of MI through time for different HbA1c thresholds for treatment initiation vs a 6.5% threshold.



Results from unweighted model (top) and Dynamic MSM (bottom) allowing one month grace period (left) and three month grace period (right). Arrow indicates upper limit of CI exceeds 2.5. CI's obtained via 200 bootstrap replications.

Using a three month grace period gave similar results for the unweighted analysis (Figure 9.7, top right and appendix 26 table 26.2). The estimates of the dynamic MSM were more similar to the unweighted analysis than for one month grace period, though, the suggested increased risk of MI with increasing HbA1c threshold in some follow up intervals was of larger magnitude than in the unweighted analysis (Figure 9.7, bottom right and appendix 26 table 26.2).

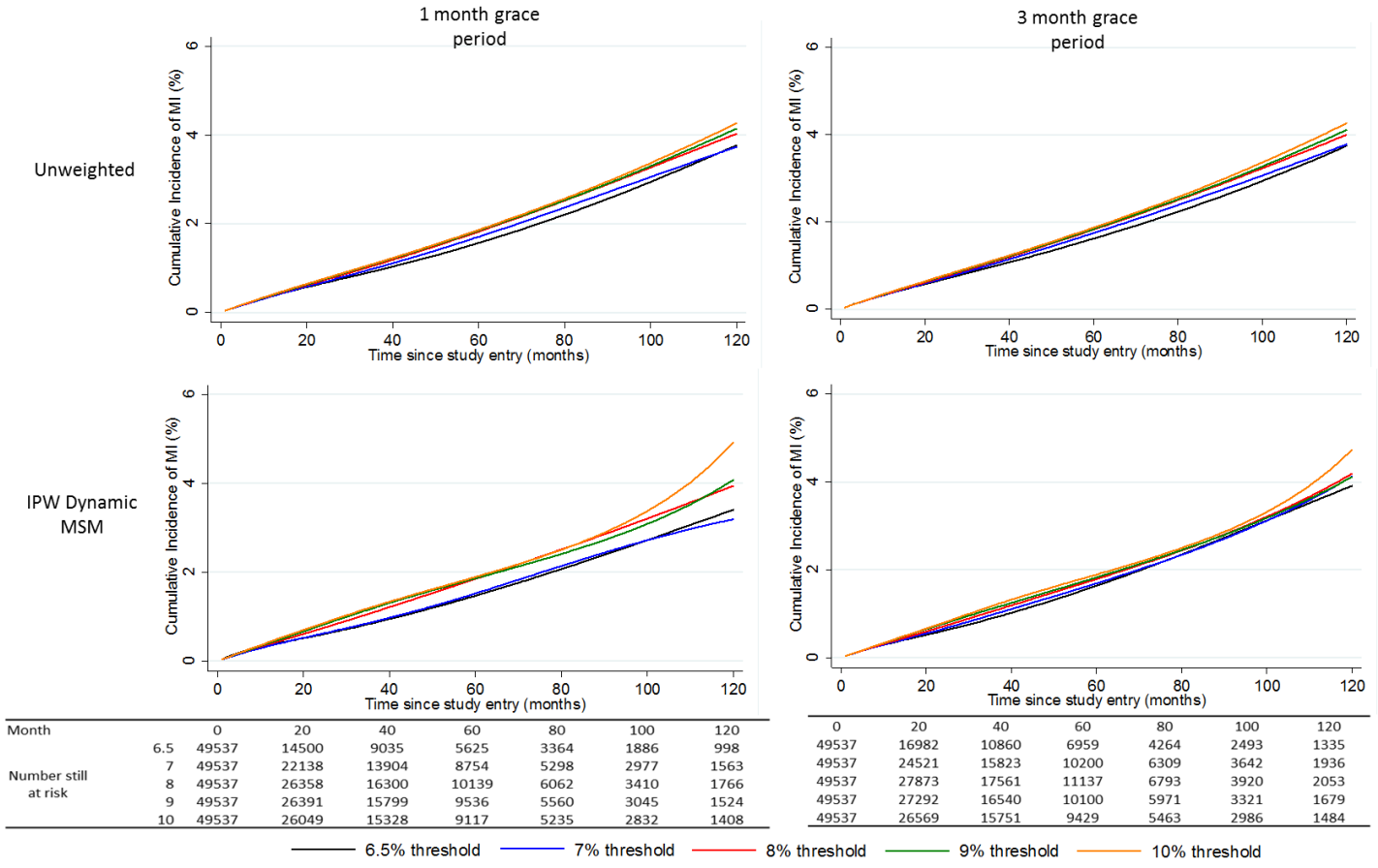
Exact estimates for cumulative incidence of MI at 1, 2 and 4 years for the model with a one month grace period are shown in Table 9.8. By 4 years, the 9% and 10 % threshold strategies had similar cumulative incidence of around 1.5%. The estimates for the 6% and 7% strategies were lower (1.13% (0.84% - 1.44%) and 1.16% (0.96% - 1.41%) respectively). There were no clear difference between the unweighted and dynamic models for these estimates of cumulative incidence. Figure 9.8 presents the cumulative incidence curves for MI, from the unweighted models (top) and the dynamic MSM (bottom) with one and three month grace periods (left and right respectively). The cumulative incidence curves for MI were similar for all strategies in the unweighted models for both grace periods. For a one month grace period the dynamic MSM showed some separation in cumulative incidence curves for strategy thresholds of 6.5 and 7%, and the 8-10% strategies from around 2 years (Figure 9.8). The 10% strategy had estimated the highest proportion of events by 10 years, though the absolute difference in cumulative incidence between the 6.5% and 10% strategy was only about 2% by 10 years, and overall the rate of MI's was low for all strategies.

		Unweighted (one-month grace period): % of population having had an MI by…			IPW Dynamic MSM (one-month grace period): % (95% CI) of population having had an MI by…		
		1 year	2 years	4 years	1 year	2 years	4 years
HbA1c threshold	6.5%	0.38	0.69	1.16	0.36 (0.23 , 0.50)	0.61 (0.46 , 0.81)	1.13 (0.85 , 1.44)
	7%	0.37	0.70	1.28	0.33 (0.25 , 0.41)	0.63 (0.51 , 0.78)	1.16 (0.96 , 1.41)
	8%	0.38	0.75	1.37	0.40 (0.32 , 0.46)	0.74 (0.63 , 0.85)	1.41 (1.17 , 1.68)
	9%	0.39	0.76	1.41	0.42 (0.34 , 0.49)	0.79 (0.66 , 0.93)	1.52 (1.25 , 1.79)
	10%	0.39	0.78	1.43	0.43 (0.36 , 0.51)	0.83 (0.69 , 0.98)	1.53 (1.31 , 1.85)

Table 9.8 Estimated cumulative incidence (%) of MI by 1, 2 and 4 years from study entry, for each treatment strategy.

95% CI given in brackets for dynamic MSM only, obtained via 200 bootstrap replications.

Figure 9.8 Cumulative incidence of MI for different HbA1c thresholds for treatment initiation.



Month		0	20	40	60	80	100	120
Number still at risk	6.5	49537	14500	9035	5625	3364	1886	998
	7	49537	22138	13904	8754	5298	2977	1563
	8	49537	26358	16300	10139	6062	3410	1766
	9	49537	26391	15799	9536	5560	3045	1524
	10	49537	26049	15328	9117	5235	2832	1408

0	20	40	60	80	100	120
49537	16982	10860	6959	4264	2493	1335
49537	24521	15823	10200	6309	3642	1936
49537	27873	17561	11137	6793	3920	2053
49537	27292	16540	10100	5971	3321	1679
49537	26569	15751	9429	5463	2986	1484

Curves are estimated from unweighted models adjusting for baseline covariates (top), and dynamic MSM with IPW (bottom). Curves are estimated allowing a one month grace period (left) or a three month grace period (right). Number at risk at time 0 excludes patients censored for death, transfer out or initiation of medication other than metformin or sulfonylureas in month 1.

9.3.4.3 Stroke

Applying a one month grace period, the dynamic MSM was suggestive of an elevated risk of stroke for all strategies vs 6.5% for 0-6 months after study entry and most prominently for 1-2 years after study entry. For time periods after 2 years, estimates tended to suggest a lower risk of stroke for higher initiation thresholds (Table 9.9 bottom, and Figure 9.9 bottom left), though the confidence intervals overlapped. Estimates from the unweighted analysis were generally consistent with no elevated stroke risk for the 7-10% thresholds compared with 6.5%, with the exception of 1-2 years, where results were suggestive of a moderate elevated risk for all strategies vs a 6.5% threshold, but the estimates for this time period were of lower magnitude than in the dynamic MSM (Table 9.9 top, Figure 9.9 top left). For the three month grace period, the results of the dynamic MSM were closer to that of the unweighted analysis (Figure 9.9, right and appendix 26, table 26.3), and showed no clear evidence of and increased or decreased stroke risk for any of the 7-10% thresholds compared to 6.5%.

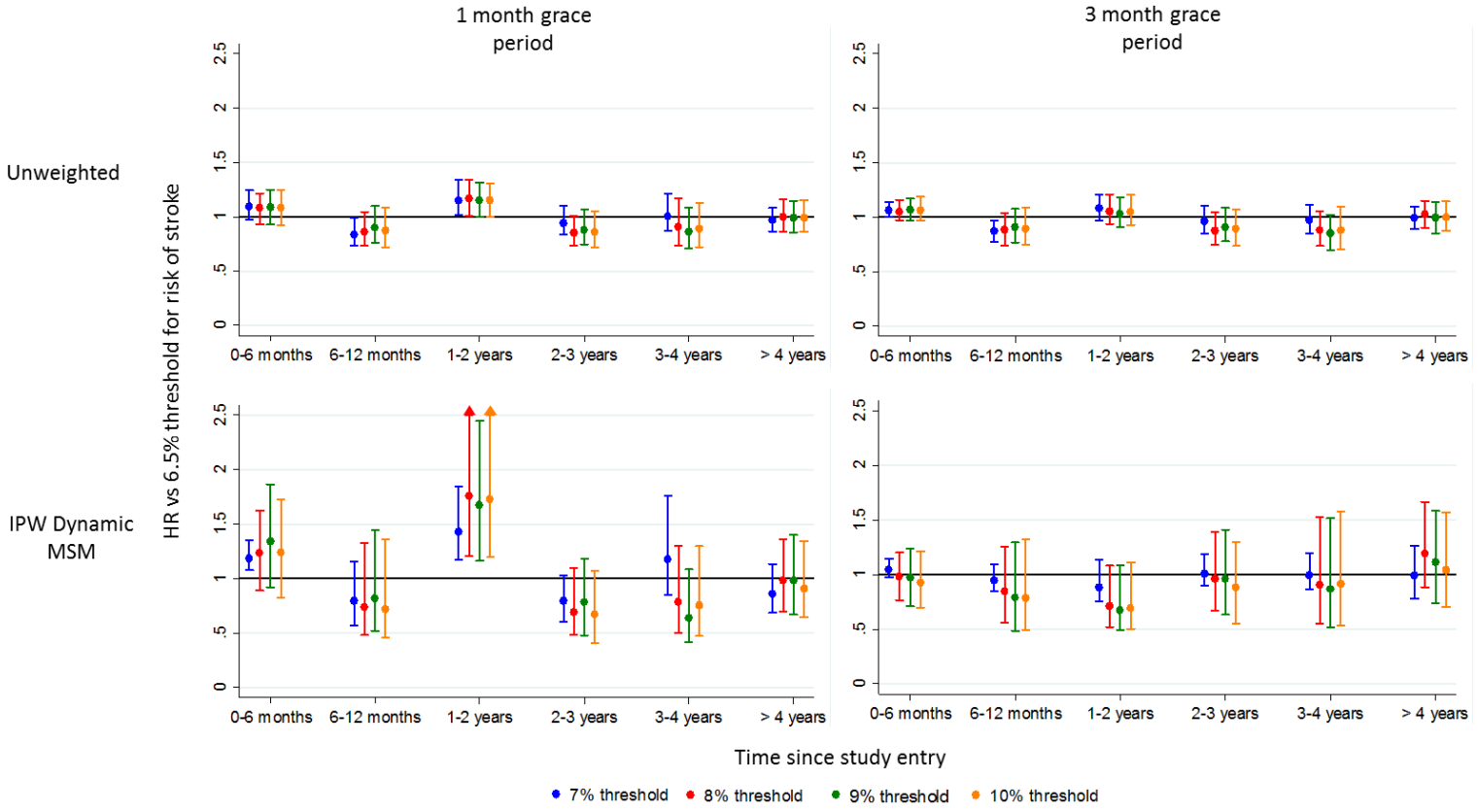
Unweighted model baseline adjusted – hazard ratio for strategy vs 6.5% for risk of stroke: one month grace period (HR<1 indicates superior strategy)						
Strategy threshold	**0-6 months[1]**	**6-12 months[1]**	**1-2 years[1]**	**2-3 years[1]**	**2-4 years[1]**	**>4 years[1]**
6.5%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
7%	1.09 (0.97 , 1.24)	0.83 (0.74 , 0.99)	1.15 (1.02 , 1.34)	0.94 (0.84 , 1.10)	1.00 (0.87 , 1.21)	0.97 (0.87 , 1.08)
8%	1.08 (0.93 , 1.22)	0.86 (0.73 , 1.04)	1.16 (1.01 , 1.34)	0.85 (0.73 , 1.01)	0.91 (0.74 , 1.17)	1.00 (0.87 , 1.16)
9%	1.09 (0.93 , 1.25)	0.90 (0.76 , 1.10)	1.15 (1.00 , 1.31)	0.88 (0.75 , 1.07)	0.86 (0.71 , 1.08)	0.99 (0.86 , 1.15)
10%	1.08 (0.92 , 1.24)	0.87 (0.72 , 1.09)	1.15 (1.00 , 1.31)	0.86 (0.72 , 1.05)	0.89 (0.72 , 1.13)	0.99 (0.86 , 1.16)
	IPW* Dynamic MSM – hazard ratio for strategy vs 6.5% for risk of stroke: one month grace period (HR<1 indicates superior strategy)					
Strategy threshold	**0-6 months[1]**	**6-12 months[1]**	**1-2 years[1]**	**2-3 years[1]**	**2-4 years[1]**	**>4 years[1]**
6.5%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
7%	1.18 (1.08 , 1.35)	0.79 (0.57 , 1.15)	1.43 (1.17 , 1.85)	0.80 (0.61 , 1.03)	1.18 (0.85 , 1.76)	0.86 (0.68 , 1.13)
8%	1.23 (0.89 , 1.62)	0.74 (0.48 , 1.32)	1.76 (1.21 , 2.56)	0.69 (0.48 , 1.09)	0.78 (0.50 , 1.31)	0.98 (0.70 , 1.36)
9%	1.34 (0.92 , 1.86)	0.82 (0.52 , 1.44)	1.67 (1.16 , 2.45)	0.78 (0.48 , 1.18)	0.64 (0.42 , 1.08)	0.98 (0.67 , 1.40)
10%	1.24 (0.82 , 1.73)	0.72 (0.46 , 1.36)	1.73 (1.19 , 2.61)	0.67 (0.41 , 1.07)	0.75 (0.47 , 1.30)	0.90 (0.65 , 1.34)

Table 9.9 Hazard ratios (and 95% CIs) to compare risk of stroke between strategy of "treat in the interval following that when HbA1c exceeds x%" for X = 7, 8, 9, 10 and reference strategy of x=6.5.

[1]`Columns show estimated HR and 95% CI by time since study entry. CI's obtained via 200 bootstrap replications

* weighting model includes: age gender, calendar period of diabetes onset (pre or post 2005), smoking status, alcohol consumption, use in the previous year of anti-hypertensive drugs, statins, NSAIDS or aspirin (baseline and time updated); previous history of any CVD, stroke, MI, CKD or cancer (baseline and time updated); HbA1c, BMI and SBP (baseline and time updated).

Figure 9.9 Estimated HRs and 95% confidence intervals to compare risk of stroke through time for different HbA1c thresholds for treatment initiation vs a 6.5% threshold.



Results from unweighted model (top) and Dynamic MSM bottom) allowing one month grace period (left) and three month grace period (right) for treatment initiation. Arrow indicates upper limit of CI exceeds 2.5. CI's obtained via 200 bootstrap replications

The estimated cumulative incidences at 1, 2 and 4 years (Table 9.10) had broadly overlapping confidence intervals for all strategies. There was some indication that the dynamic MSM estimated slightly lower cumulative incidence at 2 and 4 years compared to the unweighted model, though differences were small. The complete estimated 10 year cumulative incidence curves for stroke are displayed in Figure 9.10. Overall there was no strong evidence of any differences between strategies. In the dynamic MSM, the 10% threshold strategy was estimated to have the lowest cumulative incidence by 10 years, and the 6.5% strategy the highest. However, the absolute difference between these was very small (Figure 9.10, bottom left). In the unweighted analysis, all estimates had similar cumulative incidence curves (Figure 9.10, top).

		Unweighted (one-month grace period): % of population having had a stroke by…			IPW Dynamic MSM (one-month grace period): % (95% CI) of population having had a stroke by…		
		1 year	**2 years**	**4 years**	**1 year**	**2 years**	**4 years**
HbA1c threshold	6.5%	0.59	1.08	2.38	0.60 (0.41 , 0.81)	0.90 (0.67 , 1.11)	2.33 (1.83 , 2.80)
	7%	0.56	1.12	2.40	0.56 (0.44 , 0.73)	0.99 (0.80 , 1.18)	2.40 (2.02 , 2.77)
	8%	0.56	1.14	2.32	0.55 (0.47 , 0.67)	1.07 (0.92 , 1.25)	2.13 (1.89 , 2.35)
	9%	0.58	1.14	2.32	0.60 (0.51 , 0.70)	1.10 (0.95 , 1.23)	2.12 (1.83 , 2.37)
	10%	0.57	1.13	2.31	0.55 (0.48 , 0.63)	1.06 (0.93 , 1.20)	2.08 (1.81 , 2.29)

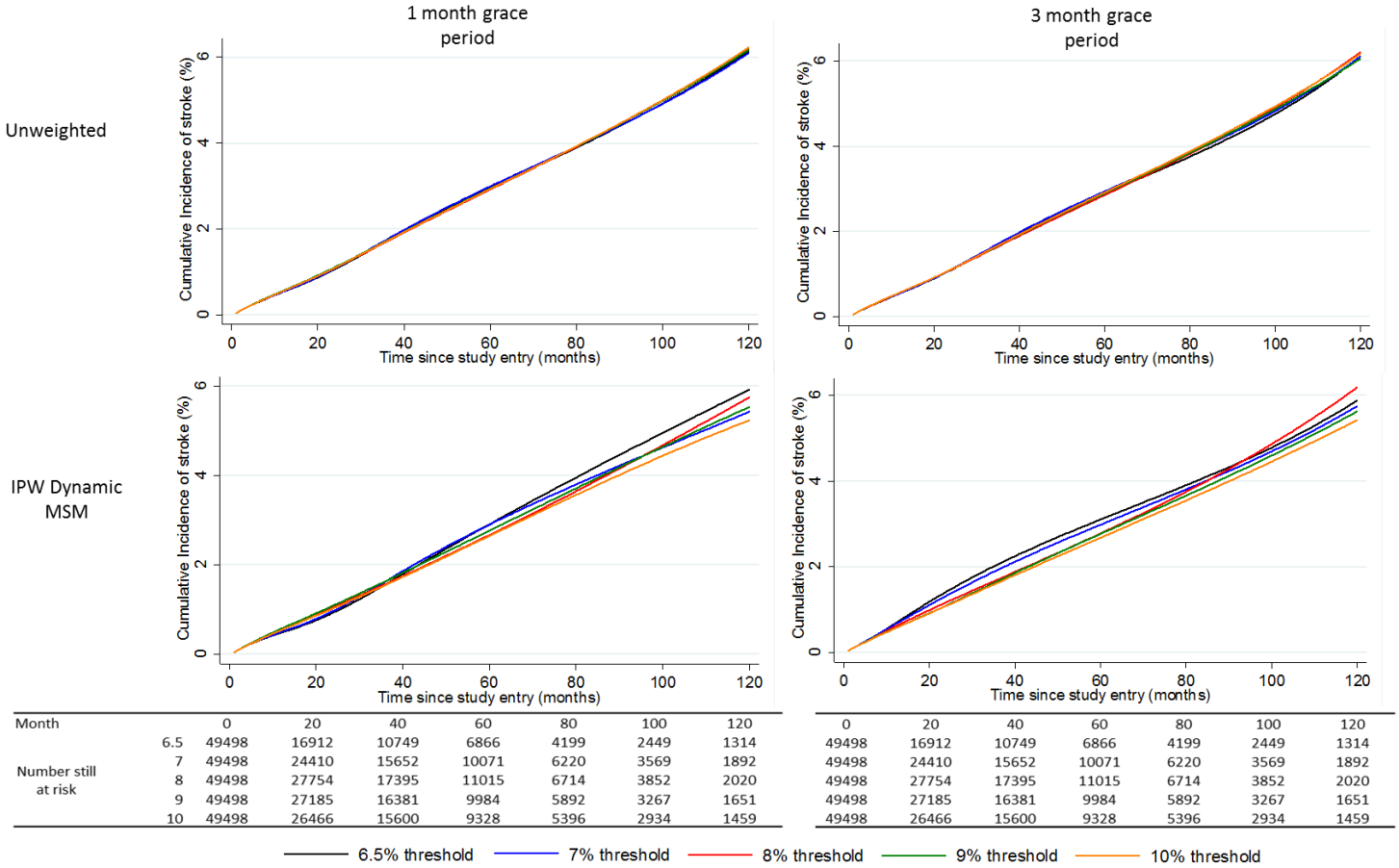
Table 9.10 Estimated cumulative incidence (%) of stroke by 1, 2 and 4 years from study entry, for each treatment strategy.

95% CI given in brackets for dynamic MSM only, obtained via 200 bootstrap replications.

9.3.4.4 All-cause mortality

All estimates for the unweighted analysis and dynamic MSM with one month grace period are given in Table 9.11, and visualised in Figure 9.11. The dynamic MSM in general estimated relative risks of all-cause mortality that were consistent with no association between treatment strategy and risk of death at any time during follow up. In general, these results were similar to those from an unweighted analysis. The main difference in the dynamic MSM was that there was no suggestion of an increased risk of mortality for any strategy in the first 6 months of follow up, whereas in the unweighted analysis, all strategies were estimated to have an increased risk of mortality in this period compared to the 6.5% threshold.

Figure 9.10 Estimated cumulative incidence of stroke for different HbA1c thresholds for treatment initiation.



Month	0	20	40	60	80	100	120
Number still at risk							
6.5	49498	16912	10749	6866	4199	2449	1314
7	49498	24410	15652	10071	6220	3569	1892
8	49498	27754	17395	11015	6714	3852	2020
9	49498	27185	16381	9984	5892	3267	1651
10	49498	26466	15600	9328	5396	2934	1459

0	20	40	60	80	100	120
49498	16912	10749	6866	4199	2449	1314
49498	24410	15652	10071	6220	3569	1892
49498	27754	17395	11015	6714	3852	2020
49498	27185	16381	9984	5892	3267	1651
49498	26466	15600	9328	5396	2934	1459

*N at risk at time 0 excludes patients censored for death, transfer out of practice or initiation of medication other than metformin or sulfonylureas in month1. Curves are estimated from unweighted models adjusting for baseline covariates (top), and dynamic MSM with IPW (bottom). Curves are estimated allowing a one month grace period for initiation (left) or a three month grace period (right)

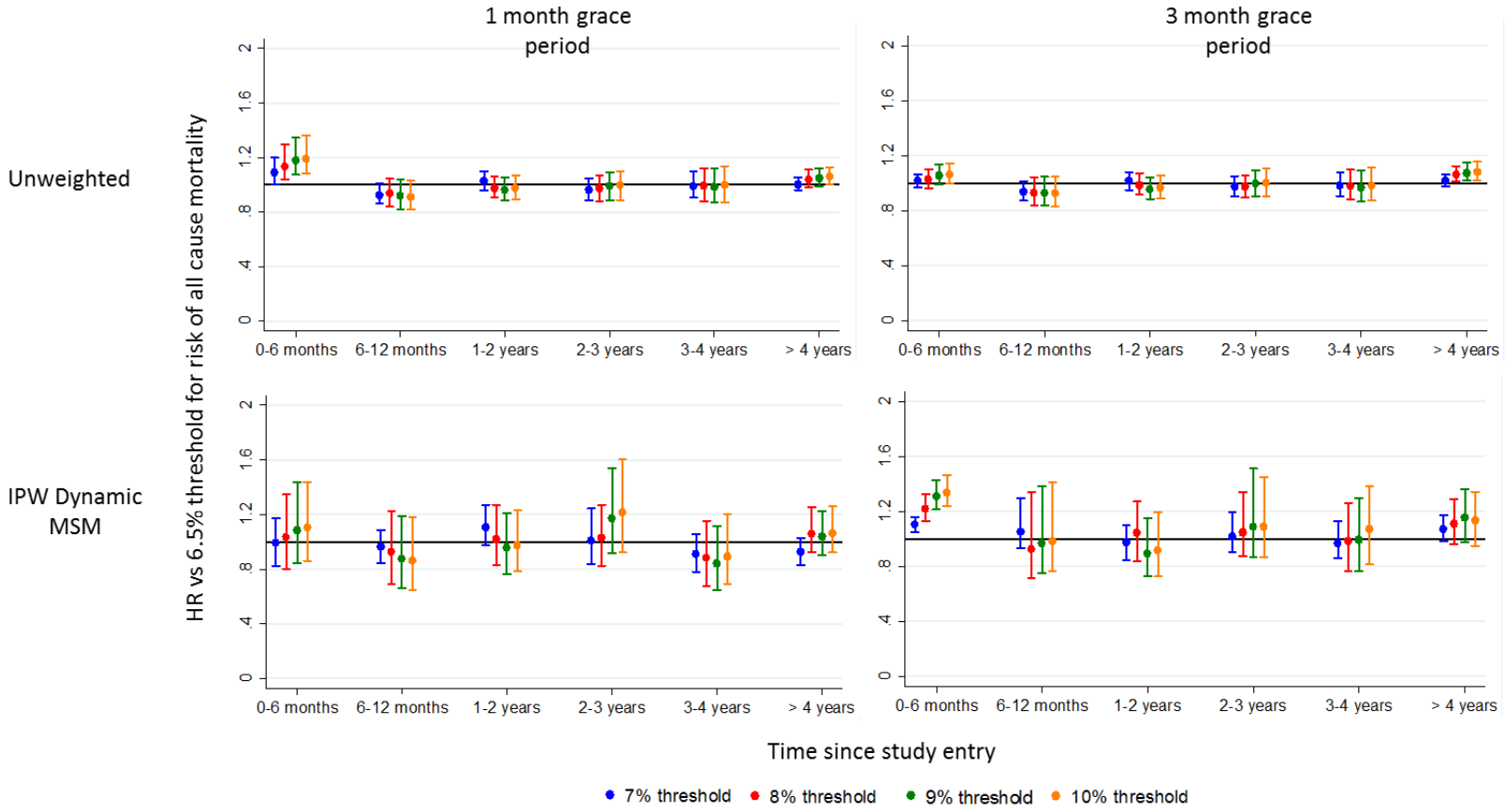
Unweighted model baseline adjusted – hazard ratio for strategy vs 6.5% for risk of all-cause mortality: one month grace period (HR<1 indicates superior strategy						
Strategy threshold	**0-6 months[1]**	**6-12 months[1]**	**1-2 years[1]**	**2-3 years[1]**	**2-4 years[1]**	**>4 years[1]**
6.5%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
7%	1.09 (1.00 , 1.2)	0.92 (0.86 , 1.01)	1.03 (0.96 , 1.10)	0.96 (0.89 , 1.05)	0.99 (0.90 , 1.10)	1.00 (0.96 , 1.05)
8%	1.13 (1.04 , 1.30)	0.93 (0.84 , 1.04)	0.97 (0.90 , 1.06)	0.97 (0.88 , 1.07)	0.99 (0.88 , 1.12)	1.04 (0.98 , 1.11)
9%	1.18 (1.07 , 1.35)	0.92 (0.82 , 1.04)	0.96 (0.88 , 1.06)	0.99 (0.89 , 1.09)	0.98 (0.87 , 1.12)	1.05 (0.99 , 1.12)
10%	1.19 (1.08 , 1.36)	0.91 (0.82 , 1.03)	0.97 (0.89 , 1.07)	0.99 (0.89 , 1.10)	1.00 (0.87 , 1.14)	1.06 (1.00 , 1.13)
	IPW* dynamic MSM – hazard ratio for strategy vs 6.5% for risk of all-cause mortality: one month grace period (HR<1 indicates superior strategy					
Strategy threshold	**0-6 months[1]**	**6-12 months[1]**	**1-2 years[1]**	**2-3 years[1]**	**2-4 years[1]**	**>4 years[1]**
6.5%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
7%	0.99 (0.82 , 1.17)	0.96 (0.84 , 1.08)	1.11 (0.97 , 1.27)	1.01 (0.84 , 1.24)	0.91 (0.78 , 1.06)	0.92 (0.83 , 1.02)
8%	1.03 (0.80 , 1.35)	0.92 (0.69 , 1.22)	1.02 (0.83 , 1.27)	1.03 (0.83 , 1.27)	0.88 (0.67 , 1.16)	1.06 (0.92 , 1.25)
9%	1.08 (0.84 , 1.44)	0.87 (0.66 , 1.19)	0.96 (0.76 , 1.21)	1.17 (0.92 , 1.54)	0.84 (0.65 , 1.12)	1.04 (0.90 , 1.22)
10%	1.10 (0.86 , 1.44)	0.86 (0.65 , 1.18)	0.97 (0.78 , 1.23)	1.22 (0.92 , 1.61)	0.89 (0.69 , 1.20)	1.06 (0.92 , 1.26)

Table 9.11 Hazard ratios (and 95% Cis) to compare risk of all-cause mortality between strategy of "treat in the interval following that when HbA1c exceeds x%" for X = 7, 8, 9, 10 and reference strategy of x=6.5. presented by time since study entry.

[1]`Columns show estimated HR and 95% CI by time since study entry. CI's obtained via 200 bootstrap replications
* weighting model includes: age gender, calendar period of diabetes onset (pre or post 2005), smoking status, alcohol consumption, use in the previous year of anti-hypertensive drugs, statins, NSAIDS or aspirin (baseline and time updated); previous history of any CVD, stroke, MI, CKD or cancer (baseline and time updated); HbA1c, BMI and SBP (baseline and time updated).

When applying a three month grace period to the dynamic MSM, the results were again more similar to the unweighted analysis, including the observation of increased risk of death for 7, 8 9 and 10% vs 6.5% in the first 6 months after study entry (Figure 9.11, bottom right and appendix 26, table 26.4).

Figure 9.11 Estimated HRs and 95% confidence intervals to compare risk of all-cause mortality through time for different HbA1c thresholds for treatment initiation vs a 6.5% threshold.



Results from unweighted model (top) and Dynamic MSM (bottom) allowing one month grace period (left) and three month grace period (right). CI's obtained via 200 bootstrap replications

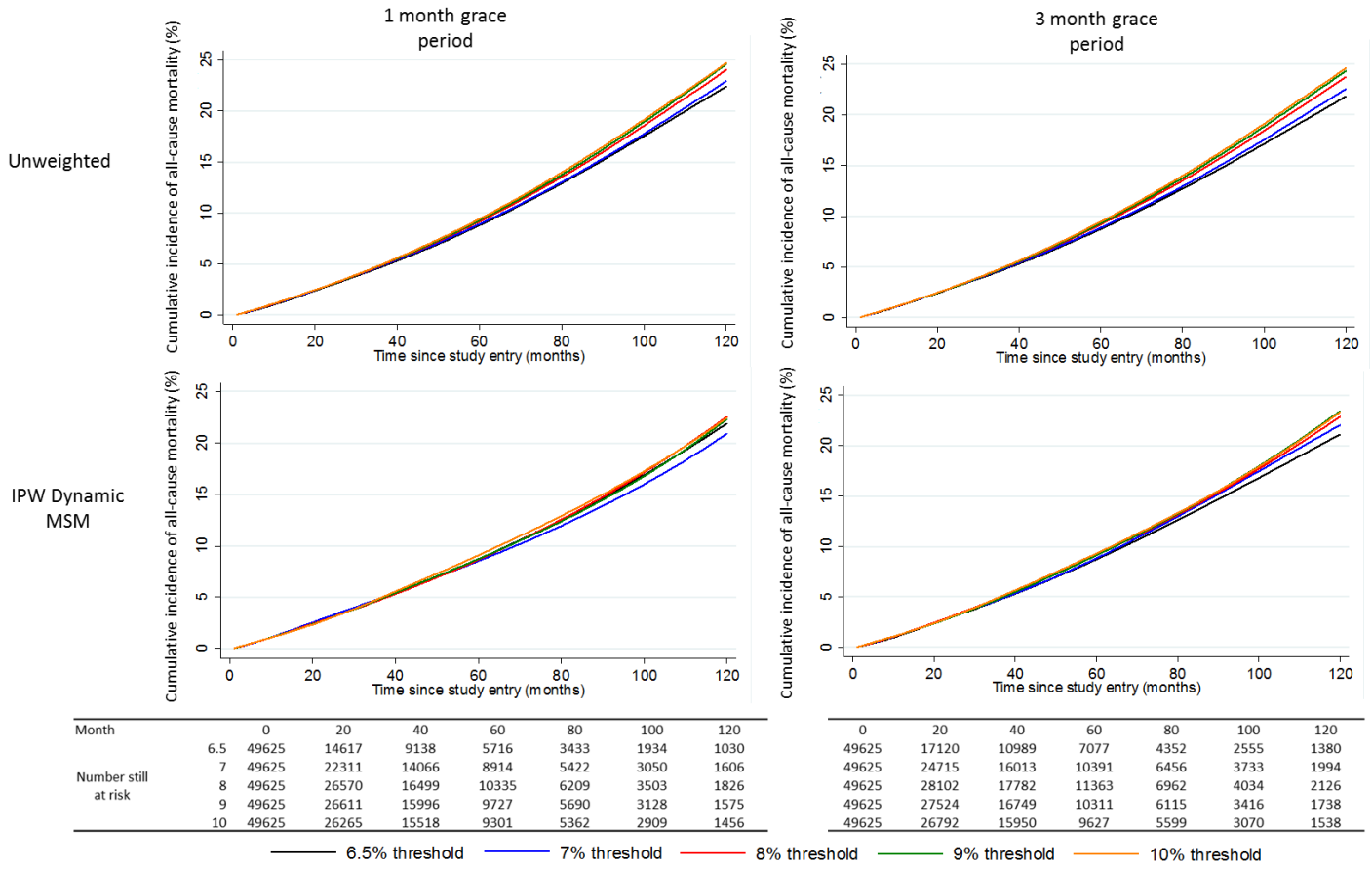
For all models, cumulative incidence curves for mortality to 10 years were very similar, and comparing exact estimates of cumulative incidence at 1, 2 and 4 years showed negligible differences between strategies in the dynamic MSM and between the dynamic MSM and the unweighted analysis (Table 9.12 and Figure 9.12).

		Unweighted (one-month grace period): % cumulative incidence of all-cause mortality at…			IPW Dynamic MSM (one-month grace period): % (95% CI) for cumulative incidence of all-cause mortality at…		
		1 year	**2 years**	**4 years**	**1 year**	**2 years**	**4 years**
HbA1c threshold	6.5%	1.39	2.94	6.75	1.53 (1.23 , 1.83)	3.04 (2.54 , 3.50)	6.97 (6.23 , 7.81)
	7%	1.43	3.06	6.85	1.49 (1.27 , 1.72)	3.16 (2.76 , 3.49)	6.90 (6.26 , 7.58)
	8%	1.48	3.05	6.98	1.49 (1.33 , 1.62)	3.02 (2.81 , 3.26)	6.73 (6.36 , 7.15)
	9%	1.49	3.05	7.03	1.47 (1.34 , 1.61)	2.92 (2.72 , 3.16)	6.77 (6.34 , 7.21)
	10%	1.49	3.07	7.10	1.47 (1.33 , 1.62)	2.94 (2.73 , 3.16)	6.97 (6.48 , 7.47)

Table 9.12 Estimated cumulative incidence (%) of all-cause mortality by 1, 2 and 4 years from study entry, for each treatment strategy.

95% CI given in brackets for dynamic MSM only, obtained via 200 bootstrap replications

Figure 9.12 Estimated cumulative incidence of all-cause mortality for different HbA1c thresholds for treatment initiation.



Curves are estimated from unweighted models adjusting for baseline covariates (top), and dynamic MSM with IPW (bottom). Curves are estimated allowing a one month grace period (left) or a three month grace period (right). N at risk at time 0 excludes patients censored for transfer out of practice or initiation of medication other than metformin or sulfonylureas in month1.

9.3.5 Sensitivity analyses

9.3.5.1 Target HbA1c of 6%

Redefining target HbA1c to be 6%, and including 6.5% as a treatment threshold in the dynamic MSM, produced broadly similar results to the primary analysis for the first two years of follow up (Table 9.13, Figure 9.13), in that strategies with higher thresholds were less likely to result in reaching target HbA1c. In contrast to the primary analysis, the hazard ratios for each strategy vs 6.5% were relatively constant up to 4 years, with some suggestion of the differences between strategies reducing beyond this time. Although the general shape of the cumulative incidence curves were similar to the primary analysis, the overall rates of target attainment were much lower (Figure 9.14). At the distinct time points of 1, 2 and 4 years, all strategies in the sensitivity analysis had cumulative incidence that estimates that were approximately half that of the same strategy in the primary analysis (Table 9.14).

IPW* dynamic MSM -hazard ratio for strategy vs 6.5% ,for reaching target HbA1c of 6%: (HR>1 indicates superior strategy)

Strategy threshold	0-6 months[1]	6-12 months[1]	1-2 years[1]	2-3 years[1]	2-4 years[1]	>4 years[1]
6.5%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
7%	0.83 (0.77 , 0.90)	0.85 (0.77 , 0.92)	0.82 (0.74 , 0.91)	0.91 (0.76 , 1.15)	0.96 (0.77 , 1.17)	1.09 (0.91 , 1.30)
8%	0.77 (0.71 , 0.85)	0.78 (0.70 , 0.88)	0.75 (0.66 , 0.85)	0.77 (0.64 , 0.97)	0.78 (0.63 , 0.97)	0.93 (0.77 , 1.16)
9%	0.75 (0.68 , 0.82)	0.76 (0.68 , 0.86)	0.72 (0.62 , 0.84)	0.72 (0.60 , 0.90)	0.77 (0.60 , 1.01)	0.97 (0.77 , 1.30)
10%	0.75 (0.68 , 0.83)	0.76 (0.68 , 0.85)	0.72 (0.62 , 0.83)	0.76 (0.60 , 1.00)	0.69 (0.52 , 0.94)	0.79 (0.62 , 1.03)

Table 9.13 Hazard ratios (and 95% CIs) to compare strategy of "treat in the interval following that when HbA1c exceeds x%" for X = 8, 9, 10 and reference strategy of x=7 in terms of reaching target HbA1c of 6%.

[1]`Columns show estimated HR and 95% CI by time since study entry. CI's obtained via 200 bootstrap replications

* weighting model includes: age gender, calendar period of diabetes onset (pre or post 2005), smoking status, alcohol consumption, use in the previous year of anti-hypertensive drugs, statins, NSAIDS or aspirin (baseline and time updated); previous history of any CVD, stroke, MI, CKD or cancer (baseline and time updated); HbA1c, BMI and SBP (baseline and time updated).

Figure 9.13 Estimated HRs and 95% confidence intervals from dynamic MSM to compare different HbA1c thresholds for treatment initiation to a 6.5% threshold, in terms of reaching target HbA1c (6%) attainment through time.

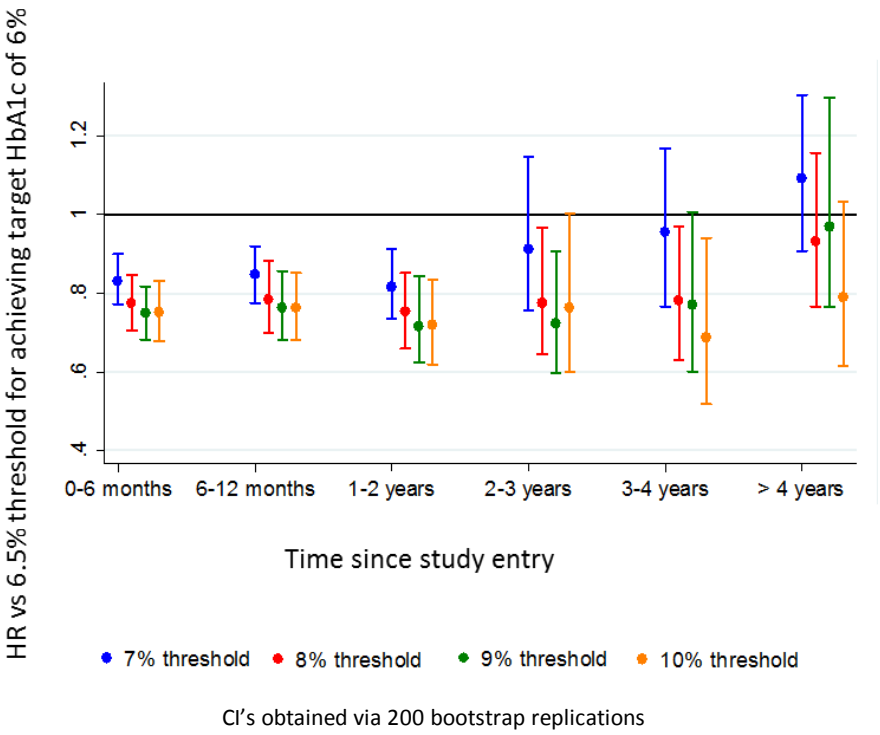
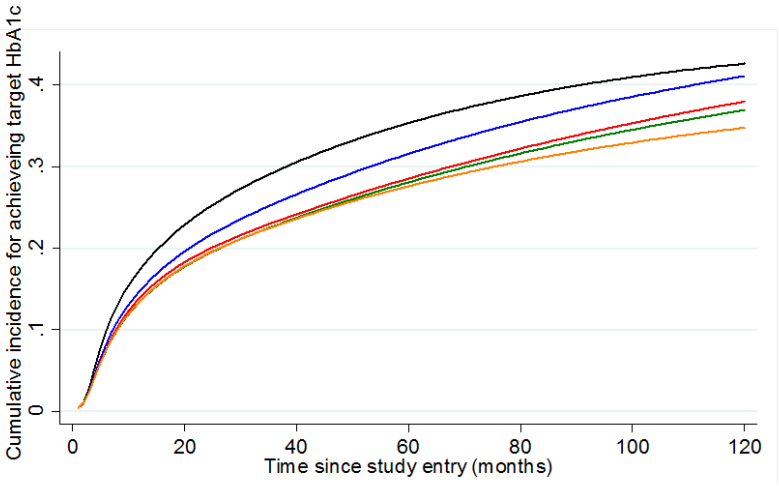


Figure 9.14 Estimated cumulative incidence curves for reaching target HbA1c of 6%, for different HbA1c thresholds for treatment initiation.



Month		0	20	40	60	80	100	120
Number still at risk	6.5	41619	5572	2814	1603	952	516	261
	7	41619	11613	5735	3065	1670	830	384
	8	41619	15275	7539	3909	2122	1077	501
	9	41619	15273	7110	3469	1754	844	365
	10	41619	14944	6711	3161	1536	706	302

		IPW Dynamic MSM **primary analysis**: proportion achieving target HBA1c of **6.5%** by…			IPW Dynamic MSM **sensitivity analysis**: proportion achieving target HBA1c of **6%** by…		
HbA1c threshold		**1 year**	**2 years**	**4 years**	**1 year**	**2 years**	**4 years**
	6.5%				0.17 (0.16 , 0.18)	0.25 (0.24 , 0.26)	0.32 (0.31 , 0.34)
	7%	0.36 (0.35 , 0.38)	0.49 (0.47 , 0.50)	0.59 (0.58 , 0.61)	0.15 (0.14 , 0.15)	0.21 (0.2 , 0.22)	0.29 (0.27 , 0.30)
	8%	0.32 (0.31 , 0.32)	0.42 (0.41 , 0.43)	0.53 (0.52 , 0.54)	0.14 (0.13 , 0.14)	0.20 (0.19 , 0.20)	0.26 (0.25 , 0.27)
	9%	0.30 (0.30 , 0.31)	0.40 (0.39 , 0.41)	0.50 (0.49 , 0.52)	0.13 (0.13 , 0.14)	0.19 (0.19 , 0.20)	0.25 (0.24 , 0.26)
	10%	0.30 (0.30 , 0.31)	0.40 (0.39 , 0.41)	0.48 (0.47 , 0.50)	0.13 (0.13 , 0.14)	0.19 (0.19 , 0.20)	0.25 (0.25 , 0.26)

Table 9.14 Estimated proportions of population achieving target HbA1c by 1, 2 and 4 years from study entry, for each treatment strategy.

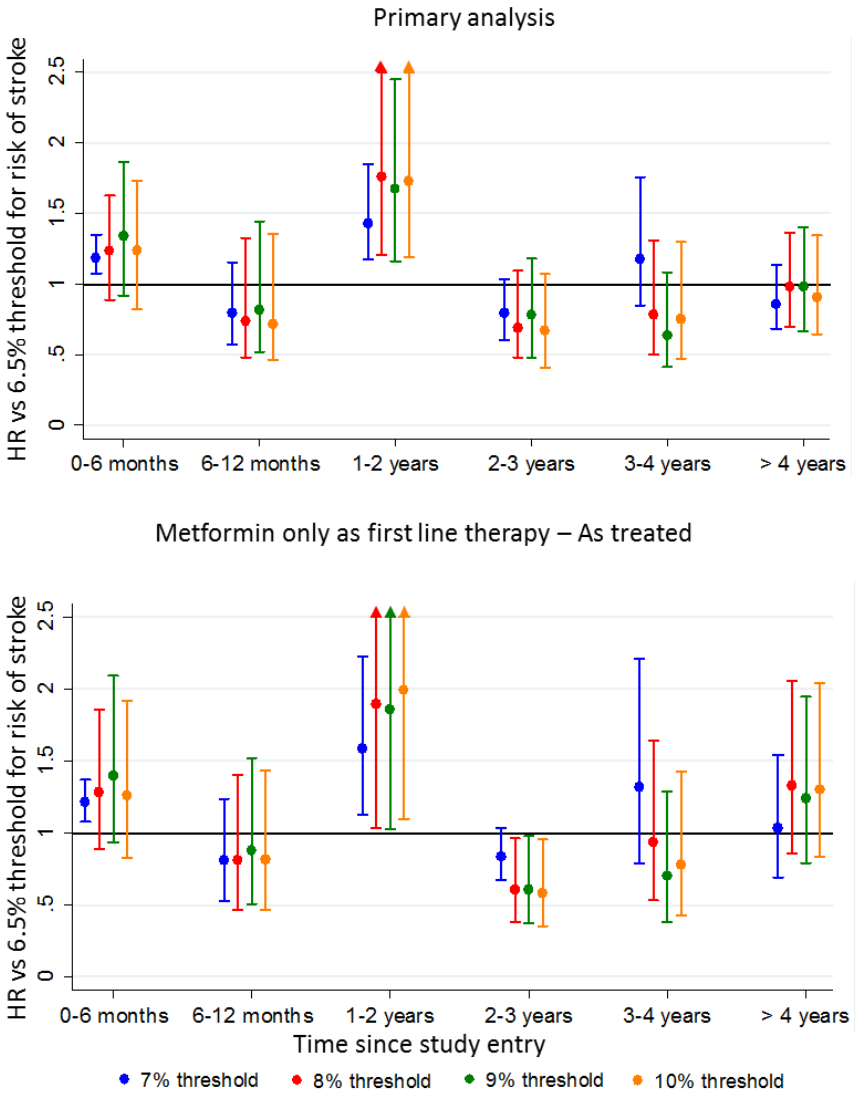
Results from primary analysis (6.5% target) left, and sensitivity analysis (target 6%) right. 95% CI's given in brackets, obtained via 200 bootstrap replications.

9.3.5.2 Initiation with metformin only

For all outcomes, the results of both the ITT style approach (treat with metformin above threshold then continue with metformin, sulfonylureas or both as necessary) and the as treated style approach (treat with metformin above threshold and continue with metformin only) to examining the effect of dynamic strategies involving metformin use alone as a first line therapy were broadly similar to the results of the primary analysis. Full results, including hazard ratio estimates, estimated cumulative incidence at 1, 2 and 4 years, and 10-year cumulative incidence curves for all outcomes are given in appendix 27.

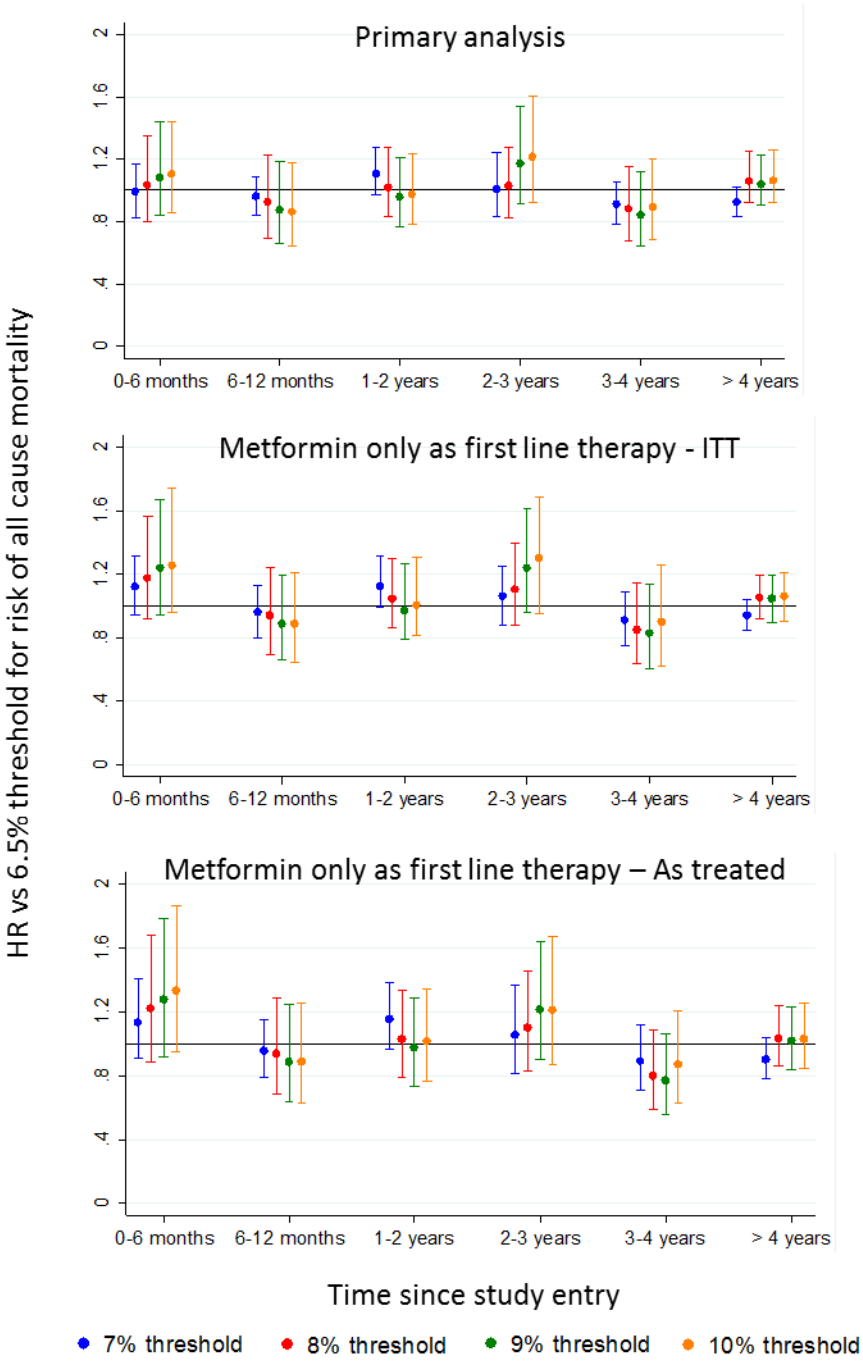
Due to greater censoring for sulfonylurea initiation in the as treated style approach, the precision of the estimates was reduced compared to the primary analysis, particularly for the HRs of the effects of strategies at 3-4 years and > 4 years after study entry. This loss of precision was most noticeable for the outcomes of MI and stroke (see appendix 27 tables 27.3 and 27.5). Although poor precision limited how clearly changes from the primary analysis could be interpreted, two notable differences were observed between the as treated sensitivity analysis and the primary analysis. Firstly, for stroke, the estimated effects of the 8, 9 and 10% threshold strategies vs a 6.5% threshold for > 4 years follow up were more consistent with an increased risk of stroke in the as treated sensitivity analysis (Figure 9.15). Secondly, for all-cause mortality, the dynamic MSM no longer reduced the estimated elevated risk of mortality for higher thresholds in the first 6-month period, which was observed in the unweighted primary analysis. The ITT and as treated approaches produced very similar results (Figure 9.16).

Figure 9.15 Estimated HRs and 95% confidence intervals to compare different HbA1c thresholds for metformin initiation to a 6.5% threshold in terms of risk of stroke through time. Results from primary analysis (top) and sensitivity analysis where patients were censored from the risk set at any initiation of sulfonylurea (bottom).



As treated style approach censors patient from the risk set at any initiation of a sulfonylurea after study entry. CI's obtained via 200 bootstrap replications

Figure 9.16 Estimated HRs and 95% confidence intervals to compare risk of all-cause mortality through time for different HbA1c thresholds of metformin initiation vs a 6.5% threshold. Results from primary analysis (top), and sensitivity analysis where patients were censored from the risk set at first line initiation of a sulfonylurea (middle) or at any initiation of a sulfonylurea (bottom).



ITT style approach censors patient from the risk set if a sulfonylurea is initiated as a first line therapy before metformin, but allows a patient to switch to or add a sulfonylurea at any time after metformin initiation. As treated style approach censors patient from the risk set at any initiation of a sulfonylurea after study entry. CI's obtained via 200 bootstrap replications

9.4 DISCUSSION

9.4.1 Main findings

The aim of this analysis was to use dynamic marginal structural models to compare dynamic treatment strategies for first line treatment initiation in patients with newly diagnosed type 2 diabetes. Four HbA1c thresholds of 7, 8, 9 and 10% were compared for how quickly a target HbA1c of 6.5% could be obtained. The same dynamic strategies and the additional strategy of a 6.5% threshold were compared for risk of MI, stroke and all-cause mortality. In estimating these effects, an additional aim was to compare the results to that of an unweighted analysis to gain insight into whether the use of dynamic MSMs appeared to adjust for expected time-dependent confounding.

For the outcome of achieving a target HbA1c of 6.5%, both the unweighted analysis and the dynamic MSM estimated that higher thresholds for initiating treatment resulted in lower incidence of attaining the target HbA1c of 6.5%. The magnitude of the differences in incidence through time was greater when estimated using the dynamic MSM, though there was loss of precision with the weighted models. Allowing for either one or three month grace periods for treatment initiation made little difference to the results.

For MI, stroke and all-cause mortality, there was no consistent evidence that different strategies had differing effects on risk of any of these outcomes through time. There was some suggestion that risk of MI was greater from 6 months onwards for more lenient strategies (i.e. higher thresholds), though for all but the 10% threshold this increase in risk appeared to diminish with time. However overall, the absolute difference between estimated incidences of MI for different strategies was small. For these outcomes, many estimates lacked precision, particularly for the weighted models, limiting how far any observed differences between strategies, or between unweighted and weighted models could be interpreted.

9.4.2 Comparison between unweighted and weighted models – interpretation and plausibility

In order to understand whether the dynamic MSMs produced plausible estimates of effect, it is important to understand the expected direction of bias in the unweighted analyses. For now, the results obtained assuming a one month grace period will be discussed. The lengthening of the grace period and its effect on the results will be discussed in section 9.4.3.

9.4.2.1 Target HbA1c

The majority of high risk patients are likely to be treated earlier, and would thus be compliant with stricter strategies dictating initiation at lower thresholds of HbA1c. Conversely, most patients who are considered lower risk might delay treatment and will therefore be compliant to the more lenient strategies. An unweighted analysis comparing strict vs lenient strategies may therefore tend to compare high vs low risk patients and so underestimate any benefit of earlier treatment. In the weighted analysis, we upweight the minority of lower risk patients who initiate treatment at lower HbA1c under the strategies with low initiation thresholds. For the more lenient strategies we upweight those who do not initiate treatment early but are at high risk, and so may be less likely to achieve target HbA1C. Therefore, the dynamic MSM would be expected to estimate a greater benefit of a lower threshold for treatment than a standard analysis. This hypothesised difference between unweighted and weighted models was observed in our analyses in both the estimated hazard ratios and the estimated cumulative incidence through time.

From a clinical perspective, the finding that delayed treatment increases the time to target HbA1c may seem obvious. If patients must wait for their HbA1c to increase further before they are treated, then it will likely take them longer to achieve target HBA1c. However, the results in Table 9.6 suggested that the proportion of on-treatment person time at higher HbA1c was greater in the strategies with the higher thresholds and therefore that the lower incidence of target achievement is not solely due to more person time off treatment as HbA1c increases; even after treatment is initiated it may take longer to reduce HbA1c from a higher level. It was suggested by some of the literature reviewed in 3.3.6, that better HbA1c control (defined by maintaining a target of 6.5%) results in lower risk of MI [126] and possibly all-cause mortality[130]. These previous findings, although not directly comparable, are not contradicted by the analyses of MI and mortality conducted here, particularly for MI (see next paragraph). Therefore, the finding here that delayed treatment may delay subsequent glucose control even after treatment initiation, could be an important result to encourage closer monitoring of patients with pre or early diabetes.

9.4.2.2 MI

MI's are potentially preventable occurrences and are not necessarily fatal. Therefore, patients presenting at high risk of MI may be treated with a more aggressive strategy in an attempt to

control their diabetes and in turn reduce risk of the event. If so, patients compliant to lenient strategies (high thresholds) may be at lower risk of events occurring. This could result in a downward bias in the estimated association between higher thresholds and MI risk in an unweighted analysis.

For 0-6 months after study entry, there was little change observed between the standard analysis and dynamic MSM. This may be expected, since risk factors such as SBP, BMI and HbA1c levels take time to change and to be updated within the CPRD. For this early time period, the unweighted model suggested a protective effect of higher thresholds for initiation, which is in line with the expected direction of confounding assuming no causal effect of strategy would be plausible within the first 6 months. However, since the estimate for 0-6 months after study entry was similar in the weighted model, there may be some residual baseline confounding.

Compared to the unweighted analysis, for 6-12 months onwards the dynamic MSM appeared to move estimated HRs in the expected direction for relative risk of MI, with higher thresholds for HbA1c in general estimated to have higher risk of MI in the dynamic MSM, but similar risk of MI in the unweighted analysis. Having said this, the magnitude of these changes was small relative to the loss in precision with the weighted analysis. For all strategies, the increase in risk in the dynamic MSM appeared to be highest at 6-12 months after study entry, with HRs observed to reduce for periods after this. One possible explanations for this pattern of risk is the varying patterns of treatment uptake within strategies. It is plausible that that more MIs occur in patients who delay treatment in the first 12 months in line with the more lenient strategies. Patients that do not have an MI in this time may then become treated in line with the particular strategy and at this point, the risk of MI reduces, since they are now treated with metformin or a sulfonylurea. In the previous chapter, the suggestion of a decreased risk of MI with metformin or sulfonylurea use was only apparent after about 4 years. This could explain why the differences to the 6.5% strategy appear to move slowly back towards the null: if those treated later at higher thresholds take time to “catch up” in terms of their reduced MI risk. Alternatively, it could be patients who do not have an MI in the first 12 months are overall at lower risk of MI therefore differences between treatment strategies are less clear after this time. Despite these possible explanations, it should be acknowledged that the low overall incidence of MIs meant the observed differences in strategies indicated by the estimated HRs differences did not translate into clinically meaningful differences in estimated cumulative incidence. For example, around a 3% cumulative incidence of MI by 10 years was estimated for the 6.5% threshold vs 5% estimated incidence for the 10% threshold.

9.4.2.3 Stroke

Overall, comparing risk of stroke between treatment initiation strategies did not produce results that were consistent with causality or that could be clearly explained, and as with MI, the magnitude of the differences between unweighted and weighted models was small when considering the width of the confidence intervals. The same direction of confounding that was postulated for MI would also be expected in the unweighted models for this outcome. However, the expected direction of change between the unweighted and weighted models was only apparent for 0-6 months and 1-2 years. For other time periods, the opposite direction of change was observed and a reduced risk of stroke was estimated for higher thresholds after 2 years of follow up. One possible explanation is suggested by the analysis in chapter 8 comparing metformin use to diet for risk of stroke, which suggested an increased risk of stroke, particularly with long term metformin use (see 8.3.4.2). Since compliance with more lenient strategies suggests less time on treatment, a drug-associated increased risk could explain the protective effect of more lenient strategies on stroke risk, which was observed in this analysis.

As discussed in the previous chapter, the reason for this increased risk of stroke was not clear, but may have been a result of residual/unmeasured confounding, ignoring intensification with sulfonylureas after metformin monotherapy in the exposure definition, or the broad definition of stroke (which may have introduced further confounding). The same issues may also have contributed to the observed results here. In the sensitivity analysis where only metformin was examined, for earlier time periods the observed results were similar to the primary analysis, but the reduction in risk for higher initiation thresholds was removed for >4 years of follow up. Also, the cumulative incidence curves were more suggestive of higher stroke incidence after 10 years with the higher initiation thresholds. However, a protective effect of a higher initiation threshold was still observed for 2-3 years in this sensitivity analysis, which cannot be clearly explained. Overall, the observed results are likely to be explained by a combination of the above factors, though the most likely explanation is some unmeasured confounding. For example, although time-varying use of statins was adjusted for, specific values of LDL cholesterol were not. If the higher perceived risk which causes the GP to treat the patients at a lower HbA1c threshold is a result of poorly controlled cholesterol, then this confounding may remain, and may be more prominent in certain periods of follow up where more stroke events occur.

9.4.2.4 All-cause mortality

As explained in the previous chapter, the expected direction of time-dependent confounding in unweighted models of all-cause mortality may not be the same as would be expected for MI and stroke. Although in patients not at immediate risk of death, the direction of expected confounding would be the same as for MI and stroke, there may also be strong confounding in the opposite direction if patients at high risk of mortality in the short term are less likely to be treated because of futility. The more lenient strategies may then look more harmful in an unweighted analysis compared to the dynamic MSM. In the unweighted analysis, a large estimated increased risk of all-cause mortality with higher HbA1c initiation thresholds was indeed observed in the first 6 months, with estimates consistent with no difference between strategies thereafter. The dynamic MSM on the other hand, did not estimate this increased risk of mortality in the first 6 months, but did not suggest evidence of differences in risk of mortality between strategies for the rest of follow up. This is consistent with the removal of confounding by frailty, but only in the early stages after study entry, which for most patients was the time of diabetes diagnosis. This could indicate an issue with reverse causality, in that the reason for mortality may also have caused the diagnosis of diabetes. However, in the sensitivity analysis where only metformin was considered, the estimates from the dynamic MSM were more similar to the unweighted analysis for 0-6 months, with an increased risk observed for higher initiation thresholds. This suggests that change in estimates between the unweighted and dynamic MSM in the primary analysis could have been driven by patients who initiate sulfonylureas soon after diabetes diagnosis. As discussed in the previous chapter in relation to the observed early increase in risk of mortality with sulfonylurea use, a possible contributing factor could be differences in severity of CKD that could not be fully adjusted for in the weighting due to positivity violations (see section 8.4.3). As previously explained, this is likely to only substantially affect the results for the outcome of mortality, since the associated risks of MI and stroke with severity of CKD could have been captured by other covariates in the model.

9.4.3 Allowing a three month grace period

For all outcomes studied, changing from a one month to a three month grace period had little impact on the results of the unweighted analyses. The one notable exception was for all-cause mortality, where the three month grace period resulted in a smaller increased risk with higher initiation thresholds in the first 6 month period. This can be explained by the fact that with a

longer grace period, those being censored from strict strategies for not being treated remained uncensored for 3 intervals instead one. Since a key reason for not being treated in line with the strategy is likely to be frailty, deaths occurring in these extra intervals would have been included for the more lenient strategies when using a three month grace period, since the censoring occurred later.

For the dynamic MSMs, the longer grace period tended to make the estimates more similar to the standard analysis. This is possibly because some of the risk factors for the outcome changed within the grace period. Although covariates such as HbA1c and BMI are unlikely to have changed substantially within a three month period, other factors such as presence of comorbidities, or use of other medications that may indicate a change in risk could quite plausibly change and be re-recorded within the space of three months. If so, since the upweighting is applied to patients complying in the final interval of the grace period only, changes in risk factors for the outcome within the grace period may result in the weighting producing a poorer balance in the risk of outcome between those complying and those not complying by the end of the grace period. The weighting approach used was one of two suggested by Cain et al [181], with an alternative being to re-weight based on the assumption of uniform initiation across the grace period. This second approach may reduce the amount of residual confounding that is re-introduced, but will not completely remove the issue if most patients still initiate in the first interval of the grace period. Since this was the case in the present analysis, it is likely that neither approach was appropriate and in fact the use of an extended grace period may have been unsuitable for this application.

9.4.4 Validity of assumptions

As alluded to in the discussion of the observed results, it is possible that unmeasured confounding could explain the inconsistent differences between standard analyses and the dynamic MSMs, and also the unexpected direction of results for some periods of follow up time. If the lack of clearly interpretable results for MI, stroke and all- cause mortality when comparing static treatment strategies in the previous chapter were a result of unmeasured confounding, then the same unmeasured confounding is likely to affect these results too.

The potential issues of model misspecification of the weighting and outcome models discussed in the previous chapter also apply to this analysis. Specific to dynamic strategies, if a difference between strategies is expected, then the proportional hazards assumption is unlikely to hold,

because for some time periods, patients will be compliant to more than one strategy, meaning the relative risk at that time between concurrent strategies must be one. It was important to make sure the model could allow for this. The interaction between time and treatment strategy was modelled in two different ways in this analysis (interactions with categorical time to estimate HRs, and an interaction with continuous time in its spline parameterisation to estimated cumulative hazard curves) to check the sensitivity of the results to misspecification of how and when the strategies may diverge. Both approaches produced results that were consistent with one another, increasing confidence that no one set of results was severely affected by misspecifying this aspect of the MSM.

In previous chapters, the extremes of the distribution of the unstabilised weights were very large, suggesting probable violations of the positivity assumption. For example, the 99th percentile of the unstabilised IPTW for the analysis of metformin and cancer risk was over 700. In contrast, the 99th percentile for the IPW for dynamic strategies was around 40, thus demonstrating the reduction in positivity issues by looking at a sensible range of treatment thresholds.

9.4.5 Other limitations

9.4.5.1 Precision

A loss of precision was observed between the weighted and unweighted analyses, which seemed to be greater than the loss when using stabilised weights in previous chapters. This loss of precision limited the extent to which the changes between the unweighted and weighted models, as well as the differences between strategies in the dynamic model, could be interpreted. The issue was most noticeable for the outcomes of MI and stroke where the incidence during follow up was lowest.

Both the weighted and unweighted confidence intervals were estimated via bootstrap replication, and it is not clear what is driving this loss in precision. Existing literature where dynamic MSMs have been applied rarely report results of unweighted analyses; therefore it is unclear whether the same phenomenon has been observed previously. In a study by Neugebauer et al [131] (previously discussed in section 3.3.6), both unweighted and weighted analyses were reported, although how the confidence intervals for the unweighted analyses were obtained was not specified. No large changes in precision of estimates between

unweighted and weighted estimates of absolute risk differences at 4 years (where the dynamic MSM CIs were based on 1000 bootstrap replications) were observed. The sample size and overall cumulative incidence of MI for the Neugebauer et al study was similar to that found in the present study, but the focus on absolute differences may have played down any differences: large differences in hazard may have little effect on the estimated cumulative incidence when cumulative incidence is low. Indeed, the precision of the cumulative incidence estimates in the present analysis appeared to be higher than for the corresponding hazard ratios.

One contributing factor to the losses in precision could be repeat sampling of subjects with extreme weights in some bootstrap replications, which may result in extreme bias in the effect estimates for those replications. With a small number of replications (200), there may be extreme results that remain even after removing the top and bottom 2.5% to obtain the confidence interval. As a check to see if this provided some explanation, the confidence intervals for the dynamic MSM looking at all-cause mortality were re-estimated using 500 replications. No notable improvements in precision were observed. Another possible explanation is that the weight estimation process is omitted from the bootstrap replications. The balancing of the censoring process with respect to the time-dependent confounders is done in the whole sample, and the weights are specific to that sample. If a particular bootstrap replication sample has, by chance, a very different distribution of covariates between those who do and don't remain compliant to particular strategies, then the weights estimated in the original sample would not be appropriate to balance the time-dependent confounding; this process could lead to different levels of confounder control in different replications, and thus an increased variability in estimates and higher empirical standard errors. This issue would not affect the unweighted model as the baseline covariate adjustment in the unweighted model is always done within the bootstrap replication. To explore this possible explanation, a simple simulation was done with a single confounder, binary exposure and binary outcome occurring in approximately 2% of the simulated sample (where the 99th percentile of the IPW for the whole sample was around 20, and truncated at this value). There were no clear differences in precision observed when including or excluding the weighting process from the bootstrap replication. However, it may be necessary to consider a more complex simulation (in particular, a simulation of a dynamic strategy), and vary the strength of confounding and the underlying incidence of the outcome, in order to establish whether the loss of precision may be explained by the reasons outlined above. This is acknowledged as a potential area for future work.

9.4.5.2 Visit frequency

For the same reasons as discussed previously (see sections 7.4.4 and 8.4.5.1), the issue of varying visit frequency and the implications for missing data may have reduced the effectiveness of the adjustment for informative noncompliance with a particular strategy. All patients must have had an HbA1c measure in order to enter the study, however if no further visits were recorded for an individual because they did not visit the GP, then depending on the value of the baseline HbA1c, the patient would have remained compliant with a set of strategies and not been censored. At the same time, potentially inaccurate information based on LOCF would have be used to upweight this individual and others when patients with what were assumed to be the same risk factors became noncompliant. For the outcome of time to target HbA1c, this individual would also never reach the target since they never had another measure, and therefore the rates of target attainment in the strategies with which these patients were apparently compliant would be underestimated. A major concern with the findings of the analysis of achieving target HbA1c, was that the decreased risk of the outcome in the most lenient strategy was driven by this. However, this would have to mean that patients with a higher HbA1c were those less likely to attend regularly. A simple check was to look at the rate of updated HbA1cs occurring for follow up on each strategy, and this was found to be similar across strategies, with observed rates of 15, 12, 12 and 13 measures per 100 person months for 7 8 9 and 10% strategies respectively. Therefore, the results of this analysis are unlikely to be entirely attributable to issues of nonattendance. It was also observed that some patients did achieve target HbA1c off treatment, and these patients must have attended to have the outcome measured. Although it was not done for this thesis, a further extension could be to explore this group further, and examine interactions between baseline covariates and strategy to see whether delaying treatment in certain subgroups may be more beneficial or detrimental for any outcomes.

9.4.5.3 Approach to dealing with diabetes treatments other than metformin or sulfonylureas

To be consistent with previous chapters, patients were treated as if lost to follow up at initiation of any diabetes treatment other than metformin or a sulfonylurea, and were censored at the beginning of the interval in which this occurred. Therefore, if for example, a patient initiated insulin after their HbA1c exceeded the given threshold for a strategy, they were censored at the beginning of the interval, before they could be counted as noncompliant for not initiation metformin or a sulfonylurea. Such a patient would have been considered to have remained

compliant to that strategy as untreated until the beginning of the interval in which insulin was initiated, and would not have contributed to the weighting model or outcome model in the interval in which they actually became noncompliant, because they were no longer considered to be at risk. If the outcome occurrence influenced the decision to initiate an alternative first line therapy, then by censoring as if lost to follow up instead of treating it as noncompliance to the strategy, some bias may have been introduced. However, it is felt that the impact of this is likely to be negligible in the present analysis, since, as reported in 6.3.1, only 2.5% of patients initiated something other than metformin or a sulfonylurea as their first line therapy.

It could also be argued, since the strategies were only focused on first line initiation, that all other treatment intensifications beyond first line initiation could have been ignored, thus including follow up after initiation of other medications, as long as it came after an initial prescription for metformin or a sulfonylurea. However, sulfonylureas are the most common 2nd line intensification to metformin, and vice versa (as reported in 6.3.1), meaning that strategies restricted to metformin and sulfonylureas may be the most clinically relevant to examine. Further, because these are the two most commonly used medications for first and second line treatment periods, the majority of 2nd line therapy follow up was captured, and most patients censored for treatment switch were censored at their third line therapy, which was only observed in 15% of the population. Also, by the time third line intensification happens, it is less plausible that differences in risk of outcomes could be directly attributed to the threshold for first line initiation, so it was considered reasonable to exclude follow up beyond this point. Further, although treatment intensification beyond these two therapies may indicate greater disease severity, or suggest contra-indications which could also affect outcome risk, in previous chapters, additional IPCW to adjust for censoring due to medication switch and other censoring events (death, transfer out of practice) made negligible differences to the estimated effects of treatment (see sections 7.3.2.5 and 8.3.4). This further justifies the opinion that censoring at initiation of therapies other than metformin and sulfonylureas is not a major limitation of this analysis.

9.4.5.4 Computational limitations

In order to compare dynamic treatment strategies, data must be expanded to create cohorts compliant to each strategy, and confidence intervals must be estimated via bootstrapping. For only 200 bootstrap replications, the time taken to run the dynamic MSM obtaining bootstrap standard errors for MI was 32 hours using a high specification PC. The long computation time

also impacted the number of regimes that it was considered practical to compare. In this context, the results of all analyses suggested that a smooth function for regime would not have been much more informative. Although it is possible that parallel computing, if available, may improve efficiency of computation, in general terms, if applying the method to other contexts where there is an a priori assumption that the effect of increasing thresholds on outcome is not linear, this could be a limitation.

9.5 CHAPTER SUMMARY

Dynamic marginal structural models were used to estimate the differences between dynamic treatment strategies for first line treatment initiation in patients with newly diagnosed type 2 diabetes. Four HbA1c thresholds of 7, 8, 9 and 10% were compared for risk of MI, stroke and all-cause mortality, as well as for achieving target HbA1c.

For MI, the dynamic MSM estimated that higher thresholds for initiation resulted in elevated risk, which appeared to peak at 6-12 months after study entry. However, the low incidence of MIs for all strategies meant the differences in strategies indicated by the estimated hazard ratios were not clearly translated to meaningful differences in cumulative incidence over 10 years. For stroke and all-cause mortality, although some changes in point estimates were observed between unweighted and weighted analyses, the dynamic MSM did not estimate any clear or consistent differences in risk of outcome between treatment strategies through time.

There was a clear trend that higher thresholds for initiation reduced the rate at which target HbA1c of 6.5% was achieved. The change between unweighted and weighted models was consistent with the expected direction of bias given the informative censoring. This result, if possible to confirm with greater precision, could have implications for treatments. Further work to examine whether this finding is the same for subgroups of diabetic patients could provide greater understanding of which patients may be more or less successful with a lifestyle change to treat early type 2 diabetes.

The addition of a grace period appeared to re-introduce confounding, particularly for the outcomes of MI, stroke and all-cause mortality. It was concluded that this was most likely a result of incorrect assumptions about initiation patterns and how risk factors for the outcome changed within the grace period.

These analyses may still be affected by the same limitations as in previous chapters, the key limitations most likely being unmeasured confounding. The strongest evidence of differing effects of strategy was observed for time to target HbA1c. As discussed in previous chapters, dynamic MSMs using data from routine practice may only be useful to examine effects of diabetes treatment strategies on outcomes that are directly targeted by or closely related to the treatment used within the strategy. Further work could examine other such outcomes to confirm this.

10 THESIS SUMMARY

This final chapter aims to provide an overall summary of the work presented. Firstly, the aims and objectives outlined at the beginning of the thesis will be recapped and the key findings relating to these aims summarised. The key strengths and limitations of the work will then be re-iterated and recommendations for future work outlined. Finally, the overall conclusions of the thesis will be presented.

10.1 RECAP OF AIMS AND OBJECTIVES

This thesis aimed to investigate potential risks and benefits of first line type 2 diabetes treatments via the application of inverse probability of treatment weighting of marginal structural models to routinely collected primary care data in the UK Clinical Practice Research Datalink (CPRD). In addition to investigating specific pharmacoepidemiological questions, a secondary aim was to establish whether the use of such methods in a complex clinical setting and data source was practical and valid.

Specific objectives under these aims were to:

a) Review the existing observational literature that examines the association between metformin use and risk of cancer.

b) Use IPTW of MSMs to estimate the effect of metformin monotherapy vs lifestyle intervention only on risk of cancer.

c) Use IPTW of MSMs to estimate the effects of metformin and sulfonylureas vs lifestyle intervention only on risks of cardiovascular events, all-cause mortality and glucose control.

d) Use existing extensions to the methodology to examine questions relating to "dynamic" treatment strategies.

e) Descriptively examine whether the anticipated issues relating to the diabetes context and routinely collected data that may affect validity of these methods appeared to be present in the data.

f) Compare the estimated effects of treatment obtained from the MSMs to those obtained via standard analysis methods, both in situations where the effect of treatment has been estimated previously in randomised trials, and where it is still unknown.

10.2 SUMMARY OF FINDINGS AND COMPARISON WITH PREVIOUS STUDIES

10.2.1 Aim 1: Apply IPTW of MSMs to investigate risk/benefits of first line diabetes therapies

10.2.1.1 Objective a: systematic review of metformin use and risk of cancer in patients with T2DM

The systematic review of metformin and risk of cancer was the first (as far as could be identified) to systematically and objectively summarise the existing research both in terms of estimated effects and potential for bias, to identify which existing studies appeared to be the most reliable. For all cancer, 18/21 estimated a protective effect of metformin on cancer risk, with 12/18 having upper confidence limits of the relative risk (either HR or OR) below one. The magnitude of the effect estimates ranged from just a 0.04% reduction in risk, to a 77% reduction in risk. For site-specific cancers estimates were also highly variable across studies. Based on the information available, only 3 of the 46 studies identified had low or no risk of bias in all domains, and none of these studies found evidence of a protective effect of metformin on risk of cancer that would be consistent with causality. Of the 12 studies that estimated a statistically significant protective effect of metformin, 11 had at least medium risk of bias in at least two domains. Nine had medium or high risk of bias from exposure definition and 7 had medium or high risk of bias for how HbA1c, BMI and other OADs were accounted for in the analysis. A meta-regression of study characteristics and bias risk could not clearly identify any factors that influenced the magnitude of estimated associations but suggested that choice of comparator and biases arising from exposure definitions may both play an important role. In an updated search, a further 23 studies were identified (see section 2.5). Across all 69 studies, none used any advanced causal methodology to deal with possible time-dependent confounding, confirming that such an analysis may be a valuable addition to the existing literature.

10.2.1.2 Objective b: Application of IPTW of MSMs to estimate the effect of metformin vs diet only on risk of cancer

Within a cohort of diabetes patients identified from the CPRD primary care database, the MSM estimated no evidence of an effect of metformin monotherapy on risk of all cancer compared to diet only, with an estimated HR of 0.97 (0.83-1.12). This finding was consistent when looking at length of exposure, with all categories of exposure having HRs within the range 0.93-1.09. A range of sensitivity analyses also gave very similar results to the main analysis. In a secondary analysis, breast, prostate, lung and pancreatic cancer were examined separately, though results were varied and inconclusive. The results of the main analysis were broadly consistent with other existing literature deemed to be at least risk of bias (see 10.2.1.1).

10.2.1.3 Objective c: Application of IPTW of MSMs to estimate the effect of metformin vs diet only and sulfonylureas vs diet only on risks of MI, stroke, all-cause mortality and glucose control

Metformin was associated with a decrease in risk of MI compared to diet only for greater than 5 years of exposure, which was consistent with the direction of the observed results of the UKPDS [31], although smaller in magnitude. There was no clear evidence that metformin was associated with a reduced risk of all-cause mortality, in contrast to the UKPDS findings and a meta-analysis of trials also discussed in chapter 3 [34]. There was some suggestion of an increased risk of stroke with long term use of metformin, which, based on the literature reviewed in chapter 3, has not been previously observed. However, this increase in risk was diminished in some sensitivity analyses. Metformin users were estimated to have an HbA1c around 0.4% lower (in absolute terms) than patients on diet only by 12 months, which although slightly smaller in magnitude than the effect estimated in the UKPDS, was consistent with a different study that considered how varying levels of metformin adherence were associated with the change in HbA1c after 1 year [223].

There were no clear patterns of increased or decreased risk through time of any of the time-to-event outcomes with sulfonylurea use vs diet only. This was broadly consistent with existing trial data comparing sulfonylureas with diet only [30, 114, 120]. The estimated early increased risk of all-cause mortality, also observed in previous observational studies [32, 33, 125] was thought to be suggestive of residual confounding rather than a true causal effect. The estimated decline in HbA1c with sulfonylurea use compared to diet only was estimated to be larger than that for metformin, with estimates ranging from a 0.5% decrease to a 1% decrease over the first 12

months, depending on model specification, which was again broadly consistent with the findings of the UKPDS given likely differences in adherence.

To our knowledge, these analyses were the first applications of IPTW of MSMs to examine the association between first line metformin and sulfonylureas vs lifestyle intervention and the risk of cardiovascular outcomes, all-cause mortality and glucose control.

10.2.1.4 Objective d: Application of dynamic marginal structural models to estimate the effect of different HbA1c thresholds for treatment initiation

Treatment strategies of the form "treat when HbA1c raises above x%" for x = 7,8,9 and 10% were compared in terms of their effect on the outcomes of reaching target HbA1c of 6.5%, MI, stroke, and all-cause mortality. x=6.5% was also included for the latter three outcomes. Overall, higher thresholds for treatment initiation were estimated to result in lower incidence of target HbA1c attainment. For example, with a threshold of 10%, 30% of patients were expected to have reached target after 1 year, with a 95% CI of 30%-31%. The corresponding figure for a threshold of 7% was 36% (35% - 38%). By 2 years, the equivalent numbers for 10 and 7% thresholds were 40% (39%-41%) and 49% (47% - 50%) respectively. There was some suggestion that delayed initiation of first line therapy increased the risk of MI, with the highest increase in risk between 6 and 12 months from study entry. However, the low overall incidence of MI observed in this study meant the estimated relative differences between strategies had limited impact on how estimated cumulative incidence varied between strategies. There was no consistent evidence of differences between strategies for risk of stroke or all-cause mortality. For all outcomes, the precision of the estimated effects was very low. No existing literature was found comparing thresholds of HbA1c for first line treatment initiation; however, Neugbauer et al [131] compared thresholds for starting second line treatments. Consistent with the analysis presented here, they found some suggestion of a non-significant trend for fewer MIs with lower treatment intensification thresholds, but the absolute differences were very small. As discussed in section 3.3.6, other studies have compared "intensive" vs "conventional" glucose control in randomised settings, with a recent large umbrella review finding no evidence that tight control reduced risk of all-cause mortality or stroke, but did observe a consistent reduction in non-fatal MI across all trials and reviews. This review also found that in all major trials [31, 114, 116, 117, 127] , HbA1c at the end of the study was lower in the intensive groups than the conventional groups. Although the question of interest in these studies is different to that addressed using the dynamic MSMs in this thesis, the findings are consistent for all outcomes examined, given

that broadly speaking, lower HbA1c thresholds for initiation would be included within the definition of "intensive" strategies, and higher thresholds within "conventional".

10.2.2 Aim 2: Investigate whether inverse probability of treatment weighting of MSMs can effectively adjust for anticipated time-dependent confounding in a complex clinical setting with a challenging data source

10.2.2.1 Objective e: Descriptive investigation of anticipated positivity violations and differing visit frequency

In an initial descriptive analysis, initiation of metformin and sulfonylureas as a first line therapy for T2DM appeared to be associated with the key confounders of age and HbA1c; and, to a smaller extent, BMI. The differences in HbA1c and BMI between those initiating the different treatments and remaining on diet only were apparent at the time of diagnosis, and as time progressed. Consistent with these initial findings, large weights were observed which increased the overall mean of the raw IPTW. Simple truncations at the 99th percentile (or more leniently at 10) were observed to move the mean close to 1. Additionally, use of simpler categorical model specifications was observed to improve the mean of the stabilised weight and did not substantially alter the estimated treatment effects. These findings suggested that simple existing methods for reducing near positivity violations were useful in this context, though the possibility that truncation of the weights re-introduced some residual confounding cannot be excluded. The distribution of the unstabilised weights from the dynamic models indicated that positivity violations were less severe in this setting, due to the selection of realistic strategies that were adhered to by at least 10% of the population.

In terms of visit frequency issues, initial investigations indicated that HbA1c and BMI were less frequently updated in patients not initiating treatment. However, there were no clear differences in number of consultations. This suggested that although the updating of measures may differ, the patients not initiating treatment were still in contact with their GP and therefore still had the opportunity to receive treatment. In this case, and under the assumption that lack of updating meant the measures remained stable (which was not contradicted by some basic descriptive analyses), it was felt that the simple approach of LOCF would be reasonable; under section 10.4.2 possible modelling of the visit process is discussed as an alternative approach.

10.2.2.2 Objective f: Comparison to standard analysis methods

For the analysis of metformin and risk of any cancer, the differences between standard analysis methods and the MSMs with both IPTW alone and joint IPTW and IPCW were minimal. The differences were greater for site specific analyses but interpretation was hampered by imprecision. For MI, stroke and all-cause mortality, there were slightly larger differences between results from MSMs and standard models, but in general, all confidence intervals were consistent between methods. Despite this, for MI and all-cause mortality, the differences observed were in the expected direction suggesting that the weighting was eliminating some likely time-dependent confounding. For longitudinal HbA1c, the standard analysis methods clearly suggested bias in the estimated treatment effect, and the MSMs with IPTW estimated the effect of treatment to be in the expected direction. For static treatment strategies, only two papers could be found that applied MSMs with IPTW using routinely collected data in the diabetes context. In the first, the authors were interested in comparing two options for intensification after metformin monotherapy in terms of risk of a composite cardiovascular outcome and found no substantial differences between standard analysis methods and use of MSMs [233]. In the second, the authors examined medication adherence in relation to risk of a composite microvascular endpoint [234], and found that the MSM with IPTW produced results that were closer to those observed previously in clinical trials compared to an unweighted analysis. These two studies support the findings in this thesis in that the differences to standard analysis methods were most pronounced for outcomes that may be more directly affected over a shorter time scale by the treatment of interest. Further, the example presented in this thesis, of using MSMs with IPTW to estimate an unbiased effect of metformin and sulfonylureas as first line therapy vs diet only on HbA1c trajectory, may be one of the first to show a clear benefit of this methodology in the diabetes context using UK primary care data.

For the dynamic models, differences between weighted and unweighted models were in the expected direction for the outcomes of target HbA1c attainment, MI and all-cause mortality. The magnitude of the differences was largest for target HbA1c attainment. The differences between models for stroke were inconsistent and not easily interpretable. For all outcomes, the magnitude of the differences between models were relatively small in relation to the loss of precision in the weighted models compared to the unweighted models (see section 10.3.2.4).

10.3 KEY STRENGTHS AND LIMITATIONS

Throughout the thesis, the limitations of each analysis have been discussed in detail in the relevant chapters. A summary of findings and main limitations for each objective are presented in Table 10.1. The focus here will be to summarise some key strengths of the research, as well as to reiterate and briefly discuss the limitations most relevant to the overall aims of the thesis.

10.3.1 Strengths

10.3.1.1 Study cohort

The analyses in this thesis were conducted within a cohort patients with newly diagnosed type 2 diabetes, that were identified using existing validated algorithms. This well-defined cohort with low risk of misclassification of diabetes status allowed for the estimation of the associations of interest in a relevant population. Additionally, the use of patients with incident type 2 diabetes eliminated any pre-existing differences in risk of any outcome that could be due to varying diabetes duration at study entry. Further, the large sample size strengthens this work since, although some analyses still lacked precision, the cohort of roughly 50,000 enabled allowed outcomes such as site specific cancers and individual cardiovascular events that would have not been feasible at all in smaller data sources such as those derived from individual diabetes clinics or prospective observational studies.

10.3.1.2 Careful methodological implementation

This thesis presents examples of careful implementation of inverse probability of treatment weighting of marginal structural models to compare both static and dynamic treatment strategies. Initial investigations were performed to ensure there were no clear violations of the positivity assumption in advance, and in all applications, the treatment and (where used) censoring models for the weights were developed in a thorough process with due consideration of the necessary assumptions, and the existing literature on weight estimation [169, 175]. In particular, covariate selection was based primarily on a-priori knowledge of potential confounders and risk factors of the outcomes of interest, and not on observed associations with treatment initiation; multiple covariate forms were used such that possible bias due to model

misspecification might be identified and investigated; and the distribution of the weights examined for each weighting model separately. In the outcome models, again, multiple covariate parameterisations were investigated to assess how much varying model specification may affect estimated associations, and treatment history modelled in multiple ways to ensure the effect of treatment on outcome was quantified appropriately.

10.3.1.3 Comprehensive sensitivity analyses

From a clinical and epidemiological perspective, this thesis is strengthened by the many sensitivity analyses that were conducted. For the analysis of metformin and risk of cancer, no sensitivity analyses changed the conclusion that there was no evidence of an effect of metformin use on risk of all cancers combined, strengthening this as a reliable conclusion that is robust to a wide range of modelling decisions (though it is acknowledged that the same modelling decisions may have had a different effect if there were a true effect of treatment). In further analyses of the effects of metformin and sulfonylureas on risks of MI, stroke and all-cause mortality, thorough sensitivity analyses not only provided confidence that the observed results were not obviously affected by most aspects of the study design that were examined, but also, where differences were observed, gave some insight into possible explanations for unexpected results. For example, the increased risk of stroke with prolonged metformin use was attenuated when patients were censored at initiation of sulfonylureas and IPCW used to adjust for informative censoring (as discussed in 8.4.3).

10.3.2 Limitations

10.3.2.1 Follow up time

The lack of power to detect long term effects of treatment due to short follow up is an acknowledged limitation of the analyses conducted. The average follow up times for the different analyses ranged from 3.8 to 4.5 years, but severe complications of diabetes may happen later than this. Additionally, any chemo-protective effect of metformin may take years to develop after initiation of treatment. Due to short average follow up in the metformin and cancer analysis, the estimated effects of long term use of metformin lacked precision. However,

the short average follow up for this analysis was partially due to censoring at any switch from metformin monotherapy, which was necessary to estimate an as treated effect vs diet only.

In later chapters, patients were allowed to remain in the risk set after switching between metformin and sulfonylureas, and an ITT estimate based on their initial therapy was calculated. This increased the amount of follow up available for analysis. For MI, stroke and mortality, less follow up may be required than cancer, particularly since patients were not required to be free from CVD at study entry. It was therefore felt that the average follow up would be sufficient to begin to detect an effect of medication in newly diagnosed T2DM. Since the effect of treatment on HBA1c trajectory (and on achieving target HbA1c) is shorter term, short follow up time is less of a limitation for these analyses.

The impact of available follow up on the ability to detect effects of medication on different outcomes may also partially explain the observation that differences between standard analysis methods and MSMs were smallest for cancer, and largest for HbA1c outcomes. It may be the case the association between confounder and outcome also develops over a long period of time, meaning that time-dependent confounding would only be problematic for longer follow up.

10.3.2.2 Unmeasured and residual confounding

One of the biggest limitations in terms of whether the estimated treatment effects can be interpreted causally, is the likelihood for unmeasured confounding. Particular examples of unmeasured confounders have been discussed previously, for example, diet and exercise. Although a large number of confounders were adjusted for, differences in risk between treatment groups may have remained. As well as confounders that were not included at all, the possibility for residual confounding by included variables remains. In particular, as has been explained, use of weight truncation to deal with near violations of positivity could re-introduce some confounding, and possible model misspecification, measurement error, or lack of detail for some confounders (e.g. smoking and alcohol) may mean not all confounding by these variables was accounted for. The strongest suggestion that the models were affected by residual confounding was in the analysis estimating associations between sulfonylureas and all-cause mortality. The models estimated an immediate increased risk within a single month of treatment initiation, which is unlikely to be plausible given the effect of treatment on HbA1c itself was not estimated to have its strongest effect until 3-6 months of exposure. It is more likely that there were factors that influence prescribing and were strongly predictive of death in the short term

that were not fully accounted for. One specific contributing factor may have been the inability to adjust for later stages of CKD because of positivity violations. To a smaller extent, there were some indications of immediate associations between treatment with metformin/sulfonylureas and MI risk, suggesting that there was also some residual confounding for this outcome.

10.3.2.3 Visit Frequency

Descriptive investigations into patient visit frequency (as summarised in 10.2.2.1) were considered important because frequency of visiting ones GP may be related to underlying health, and this is one of the key reasons for which using primary care data may have led to biases when implementing MSMs with IPTW. Although patterns of visit frequency and how this differed between treated and untreated individuals was not investigated beyond descriptive analyses, the ability of the weighted model to estimate a glucose lowering effect of metformin and sulfonylureas of a plausible magnitude when the standard analysis could not, gave some reassurance that any differences in GP contact and/or covariate measurement between treated and untreated individuals were not introducing serious bias. GPs are incentivised to encourage regular visits for patients with diabetes, which is likely to improve visits and recording, but even so, the possibility that varying visit frequency dependent upon exposure status has prevented full adjustment for time-dependent confounding cannot be excluded. For applications of IPTW of MSMs to other situations where comparisons may be made to healthier individuals who rarely visit the GP, associations between exposure status and visit frequency could be more of a problem and should be investigated descriptively prior to any analysis.

10.3.2.4 Lack of precision

An important observation throughout this thesis has been that the precision of estimates from weighted models has tended to be relatively low for time to event outcomes, particularly those where fewer events were observed, such as the site specific cancer analyses, and when estimating the effect of long term exposure. In the analysis of the static MSMs, since the weights induce non-independence, robust standard errors, which are likely to be conservative, were used. Coupled with the small number of events in site specific cancer analyses, or lower overall sample size when looking at longer term exposure, it is reasonable to have expected some loss of precision between unweighted and weighted analyses. For the dynamic strategy comparisons, the loss of precision in the estimated hazard ratios between the unweighted and

weighted models was more substantial than for the static treatment models. Both weighted and unweighted models used bootstrapping for confidence interval estimation, so an explanation for the differences in precision between models was not clear. As previously discussed in section 9.4.5.1, some simple investigations were undertaken to see whether the low precision in the dynamic MSM may have been a result of a small number of extreme estimates (resulting from repeated sampling of individuals with large weights) remaining within the percentile confidence interval because there were too few bootstrap replications, or whether the omission of the weight estimation from the bootstrapping process may have led overall to more variable confounder control across replications and thus inflated the bootstrap standard errors compared to the unweighted analysis. Neither of these investigations proved conclusive in explaining the observed loss in precision in many of the dynamic MSMs, though future work could examine more complex simulations in order to investigate this more comprehensively.

10.4 POSSIBILITIES FOR FUTURE WORK

10.4.1 Epidemiological extensions

There are various ways in which the analyses conducted within this thesis could be enhanced and extended to provide additional insight into the causal effects of first line diabetes treatments. Firstly, additional data obtained via linkage to registries, ONS records or secondary care may improve outcome detection [231], allow more specific outcomes to be examined (e.g. cardiovascular mortality), and may provide more detail or additional information on potential confounders, for example, socio-economic status from the ONS, or severity of CVD based on type and number hospital admissions. All these aspects may improve the ability of the weighting to adjust for time-dependent confounding and allow estimation of valid causal effects of treatment. Another aspect which may be interesting to examine is actual exposure to medication in terms of both dosage and adherence. All analyses in this thesis assumed that patients remained on treatment once initiated (until there was evidence of a switch), but did not take into account the continuity of or number of prescriptions relative to the prescribed daily dosage or indeed the prescribed dosage itself. It would be useful to investigate whether accounting for adherence to medication would alter any of the estimated effects of treatment. For example, a time-varying exposure based on actual number of prescriptions could be used. Alternatively, patients could be censored if their time-varying fill rate drops below a certain

level, and additional IPCW used to adjust for this, thereby estimating the effect of maintaining a particular level of adherence to treatment through time.

For target HbA1c and MI, the results of the dynamic models suggested that there may be some differences between treatment thresholds. Although tentative due to poor precision and other acknowledged limitations, these results may warrant further investigation. Specifically, it may be interesting to examine whether particular subsets of the population have better or worse outcomes on different strategies, as this may allow hypotheses to be generated regarding stratified diabetes management.

10.4.2 Methodological extensions

Methodologically, there are four key areas for future work. Firstly, it would be interesting to look at the relative performance of alternative methods for dealing with time-dependent confounding in EHR data. For example, the use of g-computation or g-estimation, which have advantages over IPTW of MSMs in that they do not require the positivity assumption to hold. A possible issue with this is the anticipated computational intensity of these methods. This could be overcome by looking at a random sub-sample of the data, but would be at the expense of precision.

Secondly, as discussed previously, the issue of differing visit frequency may be particularly important in other contexts where patients may have better overall health and no need to regularly visit the GP. In these situations, methods for modelling visit frequency may be required to reduce the potential for bias. For example, the visit process could be considered an additional "treatment", and weights calculated assuming a joint treatment process [216]. Before this could be done however, further investigations may be needed into the reasons for non-attendance vs the reason for having missing data on specific variables, as modelling the visit process would still require the need for all covariates to be measured if a visit occurs.

Third, the analyses conducted here have only been able to show a benefit of using MSMs in the situation where the effect of treatment on outcome was expected to be direct and relatively short term. Further outcomes where this may be the case, such as diabetic neuropathy, retinopathy, could be considered to investigate this further.

Finally, further work could include investigations into dealing with the wider issue of missing data. In this thesis, missing baseline and longitudinal data has been dealt with by delaying study

entry until complete data occurs, and then using LOCF. This is a pragmatic solution, but as far as could be found, there is no research into whether this is a reasonable approach. Particularly, the need for patients to be untreated at study entry means that the delayed study entry will systematically exclude patients who initiate treatment prior to having complete data, and may therefore cause bias in the treatment effect. Aside from this, the treatment of missing data is underdeveloped for MSMs, and this data set and motivating example could be useful to inform simulation work to look at a) the appropriateness of the approach used here; and b) the use of multiple imputation within the MSM framework.

10.5 Overall conclusions

Implementing MSMs with IPTW in routinely collected data to investigate the risks and benefits of first line therapy for T2DM proved to be practically achievable. Despite the acknowledged limitations, the analysis of metformin and cancer provides additional evidence that previously reported strong protective effects of metformin on cancer risk are unlikely to be causal, and adds a new perspective in terms of methodology to the existing literature. Specifically, the lack of evidence of an association between metformin and all cancers combined, and only small differences between MSMs and standard analysis methods, suggest that the variation in observed effects from previous studies was unlikely to be purely a consequence of time-dependent confounding. The conclusion that time-dependent confounding was not a major issue in the metformin/cancer setting assumes that the MSMs as implemented were capable of effectively controlling for such confounding; this seems to be the case since differences between MSMs and standard models were observed, in the expected direction, for outcomes such as MI and all-cause mortality, where stronger time-dependent confounding was expected and similar weighting models were used.

By examining HbA1c as a continuous repeated measures outcome, it was shown that IPTW of MSMs can be used to recover a known effect of treatment in the presence of strong time-dependent confounding. This gives confidence that good weighting models can be obtained from routinely collected primary care data, and that there is potential value in using such methods in these complex data sources. However, the specific situation where a clear benefit of the method was observed was one in which the effect of treatment was relatively fast, the treatment directly affected the outcome, and time-dependent confounding was known to be strong. In such situations, the MSM approach may be useful, for example, to confirm results of

clinical trials in larger/different populations, or to generate hypotheses about the effects of dynamic treatment strategies that have not been examined in trials. Less clear from this thesis are the benefits for studying longer term outcomes that may have more complex associations with both treatment and confounders. In such situations there may be more opportunity for more unmeasured confounding, and there may be insufficient follow up available in the EHR data to observe both the hypothesised confounding and the effect of treatment. For such outcomes, it may be informative to use MSMs alongside standard methods, but more research is needed, with longer follow up and across a range of clinical contexts, to establish whether MSMs can produce valid estimates of long term treatment effects using routinely collected data.

Aim	*Objective*	*Key findings*	*Key limitations*	*further work*
1.	Systematic review and bias evaluation of the effect of metformin on cancer risk.	• **Chapter 2** • Only 3/46 studies had low/no risk of bias in all domains. • Based on information available, studies with lowest risk of bias appeared consistent with no effect of treatment on risk of cancer. • An updated search found a further 23 studies - reflecting the interest in the question in recent years. • No studies identified used MSMs to examine this association.	• Only one database used, possible some papers missed. • Bias evaluation aimed to be objective but difficult to remove all subjectivity. Authors not contacted so bias evaluation only based on available information in papers	
	Application of IPTW of MSMs to estimate the effect of metformin vs diet only on risk of cancer	• **Chapter 7** • MSM results were consistent with no overall effect of metformin on cancer risk • This finding was consistent across model specifications and when considering cumulative medication use. • A range of sensitivity analyses gave similar results to the main analysis. • Results for sulfonylureas also consistent with no overall effect. • Analysis of site-specific cancer outcomes inconclusive. .	• Short follow up time • Cannot exclude residual confounding. • No linkage to cancer registry or HES therefore may have missed some events. • Too few events leading to low precision for site-specific cancer outcomes.	• Consider adherence/dosage/actual exposure in some way. E.g. look at actual number of prescriptions to estimate cumulative use. • Investigate use of multiple imputation to deal with missing longitudinal or baseline data. • Incorporate database linkages to better capture outcomes and covariates.
	Application of IPTW of MSMs to estimate the effect of metformin vs diet only and sulfonylureas vs diet only on risks of MI, stroke, all-cause mortality and glucose control. .	• **Chapter 8** • **Metformin** • Suggestion that long-term use resulted in lower risk of MI. • No evidence of reduced risk of stroke or all-cause mortality. . • Estimated to reduce HbA1c by around 0.4 - 0.5% vs diet only by 12 months. • HbA1c remained lower in treated group vs. diet only for rest of follow up but relative difference reduced. • **Sulfonylurea** • No consistent evidence of increased or decreased risk through time for any time-to-event outcome (MI, stroke, all-cause mortality). • Estimated early increased risk of all-cause mortality with sulfonylurea use, but based on low event numbers and not consistent with causality. • HbA1c estimated to be between 0.5-1% lower after 12 months compared to diet only.	• Strong suggestion of residual confounding for time-to-event outcomes during early exposure • Results for metformin and risk of stroke difficult to interpret. Possible issues with combining stroke types. • Follow up time short, small number of MI and stroke events observed • When looking at HbA1c trajectory, frequency of updating possibly limiting how well the effect can be measured for early exposure. • Estimates lacked precision for time-to-event outcomes, particularly for long-term medication use.	
	Application of dynamic MSMs to estimate the effect of different HbA1c threshold for treatment initiation.	• **Chapter 9** • Higher thresholds for initiating treatment resulted in lower incidence of attaining the target HbA1c of 6.5%. • Some evidence that higher thresholds increased risk of MI. • No consistent evidence that different strategies had differing effects on risk of stroke or all-cause mortality through time. • Addition of grace period was found not be appropriate, most likely due to risk factors changing within the period.	• Large loss of precision with weighting • Computational intensity with a large data set and monthly intervals • Some suggestions of residual confounding, • Follow up time short for MI and stroke with small number of events observed.	• Further work listed above for standard MSMs also apply to dynamic setting. • Look at HBA1c trajectory as well as target HbA1c. • Examine interactions between baseline covariates and strategy; to inform possible approaches for personalised medicine.

Table 10.1 Summary of work to achieve thesis aims and objectives, key findings, limitations, and possibilities for future work.

Aim	*Objective*	*Key findings*	*Key limitations*	*further work*
2	Descriptive investigations of anticipated positivity violations and differing visit frequency.	**Chapter 6** • Initiation of metformin and sulfonylurea monotherapy strongly associated with HbA1c and age, there were no complete violations of the positivity assumption. • The number of updated HbA1c and BMI measures was lower in non-initiators. • However, comparison of consultation rates suggested that patients not initiating treatment were visiting the GP at roughly the same rate as initiators at various points through follow up. **Chapters 7, 8 and 9** • Large weights were observed as expected. • Simple truncation at either 99th percentile (or more leniently a value of 10) reduced mean of stabilised weights to close to 1. • Categorical models had slightly smaller means for weights suggesting that coarser parameterisation reduced positivity violations. • Dynamic MSM had much smaller extremes of unstabilised weights due to selection of a relevant range of strategies for comparison.	**Chapter 6** • Only examined descriptively, and only a small selection of possible confounders examined. • Unable to tell from analysis performed the reason for differences between consultation frequency and updating of measurements. **Chapters 7, 8 and 9** • Unable to rule out that differences in visit frequency cause bias/residual confounding. • Weight truncation and categorisation may have reintroduced residual confounding. • Positivity violation due to metformin contraindication in severe CKD likely to have impacted observed results for all-cause mortality in particular.	• Further investigations to examine possible reasons for non-updating in non-initiators. E.g. look at details of consultations when measures are and are not updated. • Alternative methods such as G-computation could be examined (probably in a smaller sample) to see whether positivity issues with Hba1c/ CKD could be overcome. • Dynamic strategy could be extended to also depend upon CKD stage.
	Comparison of weighted vs unweighted models.	**Chapters 7 and 8** • For metformin and risk of any cancer, differences between standard analysis and MSM were minimal. • Differences for site-specific cancer analyses larger but limited precision made changes difficult to interpret. • For MI, stroke and all-cause mortality, slightly larger differences between methods but conclusions limited by low precision. • For MI and all-cause mortality, differences were generally in the expected direction. • For HbA1c, standard methods clearly showed bias in estimated treatment effect. MSM moved estimates in the direction of the expected effect, but estimated magnitude of change in HbA1c was slightly lower than that estimated by UKPDS, but still plausible given likely differences in adherence. **Chapter 9** • Differences between unweighted and weighted models were broadly in expected direction for MI and all-cause mortality, but magnitude of differences was small. • Differences for stroke were unclear, and not easily interpretable. • Differences for target HbA1c were also in expected direction, and were larger in magnitude than for CV outcomes.	• For time-to-event outcomes in both static and dynamic MSMs, low precision made changes between models difficult to clearly interpret. • Possible issues of adherence in real life setting being less than in an RCT, so observational and trial results may not be completely comparable.	• Look at a broader range of outcomes. • Investigate methods to model updating of measures. E.g. additional weighting for visit attendance.

Table 10.1 continued: Summary of work to achieve thesis aims and objectives, key findings, limitations and possibilities for future work.

1. Diabetes UK *What is Diabetes*. 2014 [cited 2014 16th October]; Available from: http://www.diabetes.org.uk/Guide-to-diabetes/What-is-diabetes/.
2. Diabetes UK, *Diabetes: Facts and Stats*. 2016: www.diabetes.org.uk.
3. Gatineau M, H.C., Holman N, Outhwaite H, Oldridge L, Christie A, Ells L., *Adult obesity and type 2 diabetes*. 2014, Public Health England: Oxford.
4. Diabetes UK *Diabetes Risk Factors*. 2017 [cited 2017 5th February]; Available from: https://www.diabetes.org.uk/Preventing-Type-2-diabetes/Diabetes-risk-factors/.
5. Pan, A., et al., *Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis.* The Lancet Diabetes & Endocrinology, 2015. **3**(12):958-967.
6. Harris, K.K., M. Zopey, and T.C. Friedman, *Metabolic effects of smoking cessation.* Nat Rev Endocrinol, 2016. **12**(5):299-308.
7. Freemantle, N., et al., *How strong is the association between abdominal obesity and the incidence of type 2 diabetes?* International Journal of Clinical Practice, 2008. **62**(9):1391-1396.
8. Chan, J.M., et al., *Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men.* Diabetes Care, 1994. **17**(9):961-969.
9. American Diabetes Association, *Economic costs of diabetes in the U.S. in 2012.* Diabetes Care, 2013. **36**(4):1033-1046.
10. Hex, N., et al., *Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs.* Diabet Med, 2012. **29**(7):855-62.
11. Diabetes UK, *The cost of diabetes*. 2014: www.diabetes.org.uk.
12. Willens, D., et al., *Interdisciplinary team care for diabetic patients by primary care physicians, advanced practice nurses, and clinical pharmacists.* Clinical Diabetes, 2011. **29**(2):60-68.
13. National Collaborating Centre for Chronic Conditions, *Type 2 diabetes in adults: Management NICE guidelines [NG28]*. 2015, National Institute for Health and Care Excellence: www.nice.org.uk.
14. Diabetes UK. *Diabetes Treatments*. 2014 [cited 2014 16th October]; Available from: http://www.diabetes.org.uk/Guide-to-diabetes/What-is-diabetes/Diabetes-treatments/.
15. British National Formulary, *BNF volume 64*. 2012: BMJ Group/Pharmaceutical Press.
16. Sharma, M., 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, 2016. **6**(1):e010210.
17. The Health and Social Care Information Centre, *National Diabetes Audit 2012-2013 Report 2: Complications and Mortality*. 2015: http://content.digital.nhs.uk.
18. Shah, A.D., et al., *Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1·9 million people.* The Lancet. Diabetes & Endocrinology, 2015. **3**(2):105-113.
19. Vigneri, P., et al., *Diabetes and cancer.* Endocrine-Related Cancer, 2009. **16**(4):1103-1123.
20. Giovannucci, E., et al., *Diabetes and cancer: a consensus report.* CA Cancer J Clin, 2010. **60**(4):207-21.
21. Baur, D.M., et al., *Type 2 diabetes mellitus and medications for type 2 diabetes mellitus are associated with risk for and mortality from cancer in a German primary care cohort.* Metabolism: Clinical & Experimental, 2011. **60**(10):1363-71.
22. Evans, J.M., et al., *Metformin and reduced risk of cancer in diabetic patients.* BMJ, 2005. **330**(7503):1304-5.
23. Hsieh, M.C., et al., *The influence of type 2 diabetes and glucose-lowering therapies on cancer risk in the Taiwanese.* Experimental Diabetes Research, 2012. **2012**:413782.
24. Ruiter, R., et al., *Lower risk of cancer in patients on metformin in comparison with those on sulfonylurea derivatives: results from a large population-based follow-up study.* Diabetes Care, 2012. **35**(1):119-24.
25. Currie, C.J., et al., *Mortality and other important diabetes-related outcomes with insulin vs other antihyperglycemic therapies in type 2 diabetes.* Journal of Clinical Endocrinology & Metabolism, 2013. **98**(2):668-77.
26. Currie, C.J., C.D. Poole, and E.A. Gale, *The influence of glucose-lowering therapies on cancer risk in type 2 diabetes.* Diabetologia, 2009. **52**(9):1766-77.
27. Libby, G., et al., *New users of metformin are at low risk of incident cancer: a cohort study among people with type 2 diabetes.* Diabetes Care, 2009. **32**(9):1620-5.

28. Tsilidis, K.K., et al., *Metformin does not affect cancer risk: A cohort study in the UK Clinical Practice Research Datalink analyzed like an initention to treat trial.* Diabetes Care, 2014. **37**(9):2522-2532.
29. Forst, T., et al., *Association of sulphonylurea treatment with all-cause and cardiovascular mortality: a systematic review and meta-analysis of observational studies.* Diab Vasc Dis Res, 2013. **10**(4):302-14.
30. Rados, D.V., et al., *The association between sulfonylurea use and all-cause and cardiovascular mortality: A meta-analysis with trial sequential analysis of randomized clinical trials.* PLOS Medicine, 2016. **13**(4):e1001992.
31. U.K.P.D.S, *Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34).* The Lancet, 1998. **352**(9131):854-865.
32. Maruthur, N.M., et al., *Diabetes medications as monotherapy or metformin-based combination therapy for type 2 diabetes: A systematic review and meta-analysis.* Annals of Internal Medicine, 2016. **164**(11):740-51.
33. Morgan, C.L., et al., *Association between first-line monotherapy with sulphonylurea versus metformin and risk of all-cause mortality and cardiovascular events: a retrospective, observational study.* Diabetes, Obesity and Metabolism, 2014. **16**(10):957-62.
34. Lamanna, C., et al., *Effect of metformin on cardiovascular events and mortality: a meta-analysis of randomized clinical trials.* Diabetes, Obesity and Metabolism, 2011. **13**(3):221-228.
35. Hernán, M.A. and J.M. Robins, *Causal Inference*. 2014, Chapman & Hall.
36. Cole, S.R., et al., *Illustrating bias due to conditioning on a collider.* International Journal of Epidemiology, 2010. **39**(2):417-420.
37. Robins, J.M., M.A. Hernán, and B. Brumback, *Marginal structural models and causal inference in epidemiology.* Epidemiology, 2000. **11**(5):550-60.
38. Robins, J., *A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect.* Mathematical Modelling, 1986. **7**(9–12):1393-1512.
39. Robins, J.M., et al., *G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients.* Epidemiology, 1992. **3**(4):319-36.
40. HIV-Causal Collaboration, et al., *The effect of combined antiretroviral therapy on the overall mortality of HIV-infected individuals.* AIDS, 2010. **24**(1):123-37.
41. Walker, A.S., et al., *Daily co-trimoxazole prophylaxis in severely immunosuppressed HIV-infected adults in Africa started on combination antiretroviral therapy: an observational analysis of the DART cohort.* The Lancet, 2010. **375**(9722):1278-1286.
42. Yang, S., et al., *Application of marginal structural models in pharmacoepidemiologic studies: a systematic review.* Pharmacoepidemiology and Drug Safety, 2014. **23**(6):560-71.
43. Farmer, R.E., et al., *Metformin and cancer in type 2 diabetes: a systematic review and comprehensive bias evaluation.* International Journal of Epidemiology, 2017. **46**(2):728-744.
44. Cohen, J., *A coefficient of agreement for nominal scales.* Educational and Psychological Measurement, 1960. **20**(1):37-46.
45. Banerjee, M., et al., *Beyond kappa: A review of interrater agreement measures.* Canadian Journal of Statistics, 1999. **27**(1):3-23.
46. Azoulay, L., et al., *Metformin and the incidence of prostate cancer in patients with type 2 diabetes.* Cancer Epidemiology, Biomarkers & Prevention, 2011. **20**(2):337-44.
47. Becker, C., et al., *Metformin and the risk of endometrial cancer: a case-control analysis.* Gynecologic Oncology, 2013. **129**(3):565-9.
48. Bodmer, M., et al., *Metformin does not alter the risk of lung cancer: a case-control analysis.* Lung Cancer, 2012. **78**(2):133-7.
49. Bodmer, M., et al., *Use of metformin is not associated with a decreased risk of colorectal cancer: a case-control analysis.* Cancer Epidemiology, Biomarkers & Prevention, 2012. **21**(2):280-6.
50. Bodmer, M., et al., *Use of antidiabetic agents and the risk of pancreatic cancer: a case-control analysis.* American Journal of Gastroenterology, 2012. **107**(4):620-6.
51. Bodmer, M., et al., *Use of metformin and the risk of ovarian cancer: a case-control analysis.* Gynecologic Oncology, 2011. **123**(2):200-4.
52. Bodmer, M., et al., *Long-term metformin use is associated with decreased risk of breast cancer.* Diabetes Care, 2010. **33**(6):1304-8.

53. Bosco, J.L., et al., *Metformin and incident breast cancer among diabetic women: a population-based case-control study in Denmark.* Cancer Epidemiology, Biomarkers & Prevention, 2011. **20**(1):101-11.
54. Chaiteerakij, R., et al., *Risk factors for intrahepatic cholangiocarcinoma: association between metformin use and reduced cancer risk.* Hepatology, 2013. **57**(2):648-55.
55. Chen, H.P., et al., *Metformin decreases hepatocellular carcinoma risk in a dose-dependent manner: population-based and in vitro studies.* Gut, 2013. **62**(4):606-15.
56. Dabrowski, M., *Glycated hemoglobin, diabetes treatment and cancer risk in type 2 diabetes. A case-control study.* Annals of Agricultural & Environmental Medicine, 2013. **20**(1):116-21.
57. Donadon, V., et al., *Metformin and reduced risk of hepatocellular carcinoma in diabetic patients with chronic liver disease.* Liver International, 2010. **30**(5):750-8.
58. Donadon, V., et al., *Glycated hemoglobin and antidiabetic strategies as risk factors for hepatocellular carcinoma.* World Journal of Gastroenterology, 2010. **16**(24):3025-32.
59. Li, D., et al., *Antidiabetic therapies affect risk of pancreatic cancer.* Gastroenterology, 2009. **137**(2):482-8.
60. Hassan, M.M., et al., *Association of diabetes duration and diabetes treatment with the risk of hepatocellular carcinoma.* Cancer, 2010. **116**(8):1938-46.
61. Margel, D., et al., *Association between metformin use and risk of prostate cancer and its grade.* Journal of the National Cancer Institute, 2013. **105**(15):1123-31.
62. Mazzone, P.J., et al., *The effect of metformin and thiazolidinedione use on lung cancer in diabetics.* BMC Cancer, 2012. **12**:410.
63. Monami, M., et al., *Metformin and cancer occurrence in insulin-treated type 2 diabetic patients.* Diabetes Care, 2011. **34**(1):129-31.
64. Monami, M., et al., *Sulphonylureas and cancer: a case-control study.* Acta Diabetologica, 2009. **46**(4):279-84.
65. Smiechowski, B.B., et al., *The use of metformin and the incidence of lung cancer in patients with type 2 diabetes.* Diabetes Care, 2013. **36**(1):124-9.
66. Wang, S.Y., et al., *Metformin and the incidence of cancer in patients with diabetes: a nested case-control study.* Diabetes Care, 2013. **36**(9):e155-6.
67. Buchs, A.E. and B.G. Silverman, *Incidence of malignancies in patients with diabetes mellitus and correlation with treatment modalities in a large Israeli health maintenance organization: a historical cohort study.* Metabolism, 2011. **60**(10):1379-85.
68. Chiu, C.C., et al., *Increased risk of gastrointestinal malignancy in patients with diabetes mellitus and correlations with anti-diabetes drugs: a nationwide population-based study in Taiwan.* Internal Medicine, 2013. **52**(9):939-46.
69. Ferrara, A., et al., *Cohort study of pioglitazone and cancer incidence in patients with diabetes.* Diabetes Care, 2011. **34**(4):923-9.
70. Geraldine, N., et al., *Relation between diabetes, metformin treatment and the occurrence of malignancies in a Belgian primary care setting.* Diabetes Research & Clinical Practice, 2012. **97**(2):331-6.
71. Hense, H.W., et al., *Cancer incidence in type 2 diabetes patients - first results from a feasibility study of the D2C cohort.* Diabetology and Metabolic Syndrome, 2011. **3**(1):15.
72. Home, P.D., et al., *Experience of malignancies with oral glucose-lowering drugs in the randomised controlled ADOPT (A Diabetes Outcome Progression Trial) and RECORD (Rosiglitazone Evaluated for Cardiovascular Outcomes and Regulation of Glycaemia in Diabetes) clinical trials.* Diabetologia, 2010. **53**(9):1838-45.
73. Lai, S.W., et al., *Risk of hepatocellular carcinoma in diabetic patients and risk reduction associated with anti-diabetic therapy: a population-based cohort study.* American Journal of Gastroenterology, 2012. **107**(1):46-52.
74. Lai, S.W., et al., *Antidiabetes drugs correlate with decreased risk of lung cancer: a population-based observation in Taiwan.* Clinical Lung Cancer, 2012. **13**(2):143-8.
75. Lee, M.S., et al., *Type 2 diabetes increases and metformin reduces total, colorectal, liver and pancreatic cancer incidences in Taiwanese: a representative population prospective cohort study of 800,000 individuals.* BMC Cancer, 2011. **11**:20.
76. Lehman, D.M., et al., *Statin use as a moderator of metformin effect on risk for prostate cancer among type 2 diabetic patients.* Diabetes Care, 2012. **35**(5):1002-7.

77. Morden, N.E., et al., *Further exploration of the relationship between insulin glargine and incident cancer: a retrospective cohort study of older Medicare patients.* Diabetes Care, 2011. **34**(9):1965-71.
78. Morgan, C.L., et al., *What next after metformin? A retrospective evaluation of the outcome of second-line, glucose-lowering therapies in people with type 2 diabetes.* Journal of Clinical Endocrinology & Metabolism, 2012. **97**(12):4605-12.
79. Neumann, A., et al., *Pioglitazone and risk of bladder cancer among diabetic patients in France: a population-based cohort study.* Diabetologia, 2012. **55**(7):1953-62.
80. Oliveria, S.A., et al., *Cancer incidence among patients treated with antidiabetic pharmacotherapy.* Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2008. **2**(1):47-57.
81. Qiu, H., et al., *Initial metformin or sulphonylurea exposure and cancer occurrence among patients with type 2 diabetes mellitus.* Diabetes, Obesity & Metabolism, 2013. **15**(4):349-57.
82. Redaniel, M.T., et al., *Associations of type 2 diabetes and diabetes treatment with breast cancer risk and mortality: a population-based cohort study among British women.* Cancer Causes & Control, 2012. **23**(11):1785-95.
83. Van Staa, T.P., et al., *Glucose-lowering agents and the patterns of risk for cancer: A study with the General Practice Research Database and secondary care data.* Diabetologia, 2012. **55**(3):654-665.
84. Yang, X., et al., *Low HDL cholesterol, metformin use, and cancer risk in type 2 diabetes: the Hong Kong Diabetes Registry.* Diabetes Care, 2011. **34**(2):375-80.
85. Greenland, S. and W.D. Finkle, *A critical look at methods for handling missing covariates in epidemiologic regression analyses.* Am J Epidemiol, 1995. **142**(12):1255-64.
86. Suissa, S. and L. Azoulay, *Metformin and the risk of cancer: time-related biases in observational studies.* Diabetes Care, 2012. **35**(12):2665-73.
87. Hernán, M.A., et al., *Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease.* Epidemiology, 2008. **19**(6):766-79.
88. Bosetti, C., et al., *Insulin and other antidiabetic drugs and hepatocellular carcinoma risk: a nested case-control study based on Italian healthcare utilization databases.* Pharmacoepidemiology and Drug Safety, 2015. **24**(7):771-8.
89. Cardel, M., et al., *Long-term use of metformin and colorectal cancer risk in type II diabetics: a population-based case-control study.* Cancer Medicine, 2014. **3**(5):1458-66.
90. Goossens, M.E., et al., *Influence of metformin intake on the risk of bladder cancer in type 2 diabetes patients.* British Journal of Clinical Pharmacology, 2015. **80**(6):1464-72.
91. Hagberg, K.W., et al., *Anti-diabetic medications and risk of primary liver cancer in persons with type II diabetes.* British Journal of Cancer, 2014. **111**(9):1710-7.
92. Kim, Y.I., et al., *Long-term metformin use reduces gastric cancer risk in type 2 diabetics without insulin treatment: a nationwide cohort study.* Alimentary Pharmacology & Therapeutics, 2014. **39**(8):854-63.
93. Ko, E.M., et al., *Metformin and the risk of endometrial cancer: a population-based cohort study.* Gynecologic Oncology, 2015. **136**(2):341-7.
94. Kong, A.P., et al., *Additive effects of blood glucose lowering drugs, statins and renin-angiotensin system blockers on all-site cancer risk in patients with type 2 diabetes.* BMC Medicine, 2014. **12**:76.
95. Kowall, B., W. Rathmann, and K. Kostev, *Are sulfonylurea and insulin therapies associated with a larger risk of cancer than metformin therapy? A retrospective database analysis.* Diabetes Care, 2015. **38**(1):59-65.
96. Kowall, B., et al., *No reduced risk of overall, colorectal, lung, breast, and prostate cancer with metformin therapy in diabetic patients: database analyses from Germany and the UK.* Pharmacoepidemiology and Drug Safety, 2015. **24**(8):865-74.
97. Lin, H.C., et al., *Effects of metformin dose on cancer risk reduction in patients with type 2 diabetes mellitus: A 6-year follow-up study.* Pharmacotherapy, 2014. **34**(1):36-45.
98. Mamtani, R., et al., *Incidence of bladder cancer in patients with type 2 diabetes treated with metformin or sulfonylureas.* Diabetes Care, 2014. **37**(7):1910-7.
99. Onitilo, A.A., et al., *Type 2 diabetes mellitus, glycemic control, and cancer risk.* European Journal of Cancer Prevention, 2014. **23**(2):134-40.

100. Sakoda, L.C., et al., *Metformin use and lung cancer risk in patients with diabetes.* Cancer Prevention Research, 2015. **8**(2):174-9.
101. Sehdev, A., et al., *Metformin for primary colorectal cancer prevention in patients with diabetes: a case-control study in a US population.* Cancer, 2015. **121**(7):1071-8.
102. Tsai, M.J., et al., *Metformin decreases lung cancer risk in diabetic patients in a dose-dependent manner.* Lung Cancer, 2014. **86**(2):137-43.
103. Tseng, C.H., *Metformin may reduce bladder cancer risk in Taiwanese patients with type 2 diabetes.* Acta Diabetologica, 2014. **51**(2):295-303.
104. Tseng, C.H., *Use of metformin and risk of kidney cancer in patients with type 2 diabetes.* European Journal of Cancer, 2016. **52**:19-25.
105. Tseng, C.H., *Metformin and endometrial cancer risk in Chinese women with type 2 diabetes mellitus in Taiwan.* Gynecologic Oncology, 2015. **138**(1):147-53.
106. Tseng, C.H., *Metformin reduces thyroid cancer risk in Taiwanese patients with type 2 diabetes.* PLoS ONE [Electronic Resource], 2014. **9**(10):e109852.
107. Tseng, C.H., *Metformin significantly reduces incident prostate cancer risk in Taiwanese men with type 2 diabetes mellitus.* European Journal of Cancer, 2014. **50**(16):2831-7.
108. Tseng, C.H., *Metformin may reduce breast cancer risk in Taiwanese women with type 2 diabetes.* Breast Cancer Research & Treatment, 2014. **145**(3):785-90.
109. Tseng, C.H., *Metformin reduces ovarian cancer risk in Taiwanese women with type 2 diabetes mellitus.* Diabetes/Metabolism Research Reviews, 2015. **31**(6):619-26.
110. Walker, E.J., et al., *Metformin use among type 2 diabetics and risk of pancreatic cancer in a clinic-based case-control study.* International Journal of Cancer, 2015. **136**(6):E646-53.
111. Yen, Y.C., et al., *Effect of metformin on the incidence of head and neck cancer in diabetics.* Head & Neck, 2015. **37**(9):1268-73.
112. Grundy, S.M., et al., *Diabetes and Cardiovascular Disease.* A Statement for Healthcare Professionals From the American Heart Association, 1999. **100**(10):1134-1146.
113. Home, P.D., et al., *Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial.* The Lancet, 2009. **373**(9681):2125-35.
114. U.K.P.D.S, *Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group.* The Lancet, 1998. **352**(9131):837-53.
115. Dormandy, J.A., et al., *Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomised controlled trial.* The Lancet. **366**(9493):1279-1289.
116. The ADVANCE Collaborative Group, *Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes.* New England Journal of Medicine, 2008. **358**(24):2560-2572.
117. The Action to Control Cardiovascular Risk in Diabetes Study Group, *Effects of intensive glucose lowering in type 2 diabetes.* New England Journal of Medicine, 2008. **358**(24):2545-2559.
118. Viberti, G., et al., *A Diabetes Outcome Progression Trial (ADOPT): An international multicenter study of the comparative efficacy of rosiglitazone, glyburide, and metformin in recently diagnosed type 2 diabetes.* Diabetes Care, 2002. **25**(10):1737-1743.
119. Pladevall, M., et al., *Cardiovascular risk associated with the use of glitazones, metformin and sufonylureas: meta-analysis of published observational studies.* BMC Cardiovascular Disorders, 2016. **16**:14.
120. Monami, M., S. Genovese, and E. Mannucci, *Cardiovascular safety of sulfonylureas: a meta-analysis of randomized clinical trials.* Diabetes, Obesity & Metabolism, 2013. **15**(10):938-53.
121. Pantalone, K.M., et al., *The risk of overall mortality in patients with Type 2 diabetes receiving different combinations of sulfonylureas and metformin: a retrospective analysis.* Diabetic Medicine, 2012. **29**(8):1029-35.
122. Monami, M., et al., *Are sulphonylureas all the same? A cohort study on cardiovascular and cancer-related mortality.* Diabetes/Metabolism Research Reviews, 2007. **23**(6):479-84.
123. Stratton, I.M., et al., *Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study.* BMJ 2000. **321**(7258):405-12.

124. Fox, C.S., et al., *Update on prevention of cardiovascular disease in adults with type 2 diabetes mellitus in light of recent evidence: A scientific statement from the american heart association and the American Diabetes Association.* Diabetes Care, 2015. **38**(9):1777-1803.
125. Hippisley-Cox, J. and C. Coupland, *Diabetes treatments and risk of heart failure, cardiovascular disease, and all cause mortality: cohort study in primary care.* BMJ, 2016. **354**:i3477.
126. Rodríguez-Gutiérrez, R. and V.M. Montori, *Glycemic control for patients with type 2 diabetes mellitus: Our evolving faith in the face of evidence.* Circulation: Cardiovascular Quality and Outcomes, 2016. **9**(5):504-512.
127. Duckworth , W., et al., *Glucose control and vascular complications in veterans with type 2 diabetes.* New England Journal of Medicine, 2009. **360**(2):129-139.
128. Ray, K.K., et al., *Effect of intensive control of glucose on cardiovascular outcomes and death in patients with diabetes mellitus: a meta-analysis of randomised controlled trials.* The Lancet, 2009. **373**(9677):1765-1772.
129. Hemmingsen, B., et al., *Intensive glycaemic control for patients with type 2 diabetes: systematic review with meta-analysis and trial sequential analysis of randomised clinical trials.* BMJ, 2011. **343**:d6898.
130. Arnold, L.W. and Z. Wang, *The HbA1c and all-cause mortality relationship in patients with type 2 diabetes is J-shaped: A meta-analysis of observational studies.* The Review of Diabetic Studies : RDS, 2014. **11**(2):138-152.
131. Neugebauer, R., et al., *Impact of specific glucose-control strategies on microvascular and macrovascular outcomes in 58,000 adults with type 2 diabetes.* Diabetes Care, 2013. **36**(11):3510-6.
132. Patorno, E., et al., *Observational studies of the association between glucose-lowering medications and cardiovascular outcomes: addressing methodological limitations.* Diabetologia, 2014. **57**(11):2237-50.
133. Ferrannini, E. and R.A. DeFronzo, *Impact of glucose-lowering drugs on cardiovascular disease in type 2 diabetes.* European Heart Journal, 2015. **36**(34):2288-96.
134. Hemmingsen, B., et al., *Comparison of metformin and insulin versus insulin alone for type 2 diabetes: systematic review of randomised clinical trials with meta-analyses and trial sequential analyses.* BMJ, 2012. **344**:e1771.
135. Neyman, J.S., D.M. Dabrowska, and T.P. Speed, *On the application of probability theory to agricultural experiments. Essay on principles. Section 9 (1990 translation of polish original).* Statistical Science, 1923. **5**(4):465-472.
136. Rubin, D., *Estimating causal effects of treatments in randomized and non randomized studies.* Journal of Educational Psychology, 1974. **66**(5):668-701.
137. Baron, R.M. and D.A. Kenny, *The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations.* Journal of Personality and Social Psychology, 1986. **51**(6):1173-82.
138. Daniel, R.M., et al., *Methods for dealing with time-dependent confounding.* Statistics in Medicine, 2013. **32**(9):1584-618.
139. Taubman, S.L., et al., *Intervening on risk factors for coronary heart disease: an application of the parametric g-formula.* International Journal of Epidemiology, 2009. **38**(6):1599-1611.
140. Daniel, R.M., B.L.D. Stavola, and S.N. Cousens, *gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula.* Stata Journal, 2011. **11**(4):479-517.
141. Keil, A.P., et al., *The parametric g-formula for time-to-event data: intuition and a worked example.* Epidemiology, 2014. **25**(6):889-97.
142. Cole, S.R., et al., *Analysis of occupational asbestos exposure and lung cancer mortality using the g formula.* American Journal of Epidemiology, 2013. **177**(9):989-96.
143. Hernán, M.A., B. Brumback, and J.M. Robins, *Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men.* Epidemiology, 2000. **11**(5):561-70.
144. Robins, J. *Marginal Structural Models*. 1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science, 1998. 1-10.
145. Fewell, Z., et al., *Controlling for time-dependent confounding using marginal structural models.* The Stata Journal, 2004. **4**(4):402-420.
146. Alexander, J.H., et al., *Apixaban vs. warfarin with concomitant aspirin in patients with atrial fibrillation: Insights from the ARISTOTLE trial.* European Heart Journal, 2014. **35**(4):224-232.

147. Crook, A.M., et al., *Injectable and oral contraceptives and risk of HIV acquisition in women: an analysis of data from the MDP301 trial.* Human Reproduction, 2014. **29**(8):1810-1817.
148. Morrison, C.S., et al., *Hormonal contraception and the risk of HIV acquisition among women in South Africa.* AIDS, 2012. **26**(4):497-504.
149. McCoy, S.I., et al., *Oral and injectable contraception use and risk of HIV acquisition among women in sub-Saharan Africa.* AIDS, 2013. **27**(6):1001-1009.
150. Rosenblum, M., et al., *Analysing direct effects in randomized trials with secondary interventions: an application to human immunodeficiency virus prevention trials.* Journal of the Royal Statistical Society Series a-Statistics in Society, 2009. **172**:443-465.
151. Vansteelandt, S. and M. Joffe, *Structural nested models and g-estimation: The partially realized promise.* Statistical Science, 2014. **29**(4):707-731.
152. Robins, J.M., *Correcting for non-compliance in randomized trials using structural nested mean models.* Communications in Statistics - Theory and Methods, 1994. **23**(8):2379-2412.
153. Hernán, M.A., et al., *Structural accelerated failure time models for survival analysis in studies with time-varying treatments.* Pharmacoepidemiology and Drug Safety, 2005. **14**(7):477-91.
154. Robins, J., *Estimation of the time-dependent accelerated failure time model in the presence of confounding factors.* Biometrika, 1992. **79**(2):321-334.
155. Picciotto, S. and A.M. Neophytou, *G-estimation of structural nested models: Recent applications in two subfields of epidemiology.* Current Epidemiology Reports, 2016. **3**(3):242-251.
156. Naimi, A.I., et al., *Estimating the effect of cumulative occupational asbestos exposure on time to lung cancer mortality: using structural nested failure-time models to account for healthy-worker survivor bias.* Epidemiology, 2014. **25**(2):246-54.
157. Keil, A.P., D.B. Richardson, and M.A. Troester, *Healthy worker survivor bias in the Colorado Plateau uranium miners cohort.* Am J Epidemiol, 2015. **181**(10):762-70.
158. Tilling, K., J.A.C. Sterne, and M. Szklo, *Estimating the effect of cardiovascular risk factors on all-cause mortality and incidence of coronary heart disease using g-estimation: The atherosclerosis risk in communities study.* American Journal of Epidemiology, 2002. **155**(8):710-718.
159. Neophytou, A.M., et al., *A structural approach to address the healthy-worker survivor effect in occupational cohorts: an application in the trucking industry cohort.* Occupational and Environmental Medicine, 2014. **71**(6):442-7.
160. Picciotto, S., et al., *Hypothetical interventions to limit metalworking fluid exposures and their effects on COPD mortality: G-estimation within a public health framework.* Epidemiology, 2014. **25**(3):436-43.
161. Sterne, J.A.C. and K. Tilling, *G-estimation of causal effects, allowing for time-varying confounding.* Stata Journal, 2002. **2**(2):164-182.
162. de Stavola, B.L., R.M. daniel, and R. Silverwood, *A short course on Concepts and Methods in Causal Inference*. 2014: Unpublished course notes.
163. Cole, S.R. and C.E. Frangakis, *The consistency statement in causal inference: A definition or an assumption?* Epidemiology, 2009. **20**(1):3-5.
164. Petersen, M.L., et al., *Diagnosing and responding to violations in the positivity assumption.* Statistical Methods in Medical Research, 2012. **21**(1):31-54.
165. StataCorp, *Stata Statistical Software: Release 14*. 2015, StataCorp LP: College Station, TX.
166. Hernán, M.A., B.A. Brumback, and J.M. Robins, *Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures.* Statistics in Medicine, 2002. **21**(12):1689-709.
167. Keogh, R., et al., *Analysis of longitudinal studies: Adjusting for time-dependent confounding using conventional methods* American Journal of Epidemiology (In Press), 2016.
168. Robins, J.M. and D.M. Finkelstein, *Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests.* Biometrics, 2000. **56**(3):779-88.
169. Cole, S.R. and M.A. Hernan, *Constructing inverse probability weights for marginal structural models.* American Journal of Epidemiology, 2008. **168**(6):656-64.
170. D'Agostino, R.B., et al., *Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study.* Statistics in Medicine, 1990. **9**(12):1501-15.
171. Gupta, S.K., *Intention-to-treat concept: A review.* Perspectives in Clinical Research, 2011. **2**(3):109-112.
172. StataCorp, *Stata 14 Base Reference Manual* 2015, Stata Press. : College Station, TX.

173. Bodnar, L.M., et al., *Marginal structural models for analyzing causal effects of time-dependent treatments: an application in perinatal epidemiology.* American Journal of Epidemiology, 2004. **159**(10):926-34.
174. Hernán, M.A. and J.M. Robins, *Estimating causal effects from epidemiological data.* Journal of Epidemiology and Community Health, 2006. **60**(7):578-586.
175. Lefebvre, G., J.A.C. Delaney, and R.W. Platt, *Impact of mis-specification of the treatment model on estimates from a marginal structural model.* Statistics in Medicine, 2008. **27**(18):3629-3642.
176. Young, J.G., et al., *Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula.* Statistics in Biosciences, 2011. **3**(1):119-143.
177. Robins, J., L. Orellana, and A. Rotnitzky, *Estimation and extrapolation of optimal treatment and testing strategies.* Statistics in Medicine, 2008. **27**(23):4678-721.
178. van der Laan, M.J., M.L. Petersen, and M.M. Joffe, *History-adjusted marginal structural models and statically-optimal dynamic treatment regimens.* The International Journal of Biostatistics, 2005. **1**(1):Article 4.
179. Orellana, L., A. Rotnitzky, and J.M. Robins, *Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: main content.* The International Journal of Biostatistics, 2010. **6**(2):Article 8.
180. Hernán, M.A., et al., *Comparison of dynamic treatment regimes via inverse probability weighting.* Basic & Clinical Pharmacolgy Toxicology, 2006. **98**(3):237-42.
181. Cain, L.E., et al., *When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data.* The International Journal of Biostatistics, 2010. **6**(2):Article 18.
182. Ewings, F.M., et al., *Optimal CD4 count for initiating HIV treatment: impact of CD4 observation frequency and grace periods, and performance of dynamic marginal structural models.* Epidemiology, 2014. **25**(2):194-202.
183. Shortreed, S.M. and E.E.M. Moodie, *Estimating the optimal dynamic antipsychotic treatment regime: Evidence from the sequential multiple assignment randomized CATIE Schizophrenia Study.* Journal of the Royal Statistical Society. Series C, Applied statistics, 2012. **61**(4):577-599.
184. Wang, L., et al., *Evaluation of viable dynamic treatment regimes in a sequentially randomized trial of advanced prostate cancer.* Journal of the American Statistical Association, 2012. **107**(498):493-508.
185. Ewings, F.M., *Practical and theoretical considerations of the application of marginal structural models to estimate the causal effects of treatment in HIV infection*, in *Medical Research Council Clinical Trials Unit*. 2010, University College London: London.
186. Carpenter, J. and J. Bithell, *Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians.* Statistics in Medicine, 2000. **19**(9):1141-64.
187. Medicines and Healthcare products Regulatory Agency, *Release Notes - CPRD GOLD January 2016*. 2016: London.
188. Herrett, E., et al., *Validation and validity of diagnoses in the General Practice Research Database: a systematic review.* British Journal of Clinical Pharmacology, 2010. **69**(1):4-14.
189. Herrett, E., et al., *Data Resource Profile: Clinical Practice Research Datalink (CPRD).* International Journal of Epidemiology, 2015. **44**(3):827-836.
190. Doran, T., et al., *Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework.* BMJ, 2011. **342**:d3590.
191. Eastwood, S.V., et al., *Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank.* PLoS ONE, 2016. **11**(9):e0162388.
192. Centers for Disease Control and Prevention. *Incidence of diagnosed diabetes per 1,000 population aged 18-79 years, by age, Unites States, 1980-2014*. 2015 17/01/2017]; Available from: https://www.cdc.gov/diabetes/statistics/incidence/fig3.htm.
193. Lewis, J.D., et al., *The relationship between time since registration and measured incidence rates in the General Practice Research Database.* Pharmacoepidemiology and Drug Safety, 2005. **14**(7):443-51.
194. Hernán, M.A., *Does water kill? A call for less casual causal inferences.* Annals of Epidemiology, 2016. **26**(10):674-680.
195. Bhaskaran, K., et al., *Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults.* The Lancet, 2014. **384**(9945):755-65.

196. Renehan, A.G., et al., *Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies.* The Lancet, 2008. **371**(9612):569-578.
197. Stengel, B., *Chronic kidney disease and cancer: a troubling connection.* Journal of Nephrology, 2010. **23**(3):253-62.
198. NHS Digital. *QOF business rules v30.0*. 2015 [cited 2015 January 18th]; Available from: http://content.digital.nhs.uk/qofbrv30.
199. Koene, R.J., et al., *Shared risk factors in cardiovascular disease and cancer.* Circulation, 2016. **133**(11):1104-1114.
200. Stocks, T., et al., *Blood pressure and risk of cancer incidence and mortality in the metabolic syndrome and cancer project.* Hypertension, 2012. **59**(4):802-810.
201. Harrell, F.E., *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. 2001, Springer: New York.
202. Royston, P. and W. Sauerbrei, *Multivariable modeling with cubic regression splines: A principled approach.* Stata Journal, 2007. **7**(1):45-70.
203. Stevens, R.J., et al., *Cancer outcomes and all-cause mortality in adults allocated to metformin: systematic review and collaborative meta-analysis of randomised clinical trials.* Diabetologia, 2012. **55**(10):2593-603.
204. Johnson, J.A. and S.L. Bowker, *Intensive glycaemic control and cancer risk in type 2 diabetes: a meta-analysis of major trials.* Diabetologia, 2011. **54**(1):25-31.
205. de Beer, J.C. and L. Liebenberg, *Does cancer risk increase with HbA1c, independent of diabetes?* British Journal of Cancer, 2014. **110**(9):2361-8.
206. Reeves, G.K., et al., *Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study*. BMJ, 2007. **335**(7630):1134.
207. Albanes, D., A. Blair, and P.R. Taylor, *Physical activity and risk of cancer in the NHANES I population.* American Journal of Public Health, 1989. **79**(6):744-750.
208. Giovannucci, E., et al., *Physical activity, obesity, and risk for colon cancer and adenoma in men.* Annals of Internal Medicine, 1995. **122**(5):327-334.
209. Thune , I., et al., *Physical activity and the risk of breast cancer.* New England Journal of Medicine, 1997. **336**(18):1269-1275.
210. Wiseman, M., *The second World Cancer Research Fund/American Institute for Cancer Research expert report. Food, nutrition, physical activity, and the prevention of cancer: a global perspective.* The Proceedings of the Nutrition Society, 2008. **67**(3):253-6.
211. Booth, H.P., A.T. Prevost, and M.C. Gulliford, *Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011.* Pharmacoepidemiol Drug Saf, 2013. **22**(12):1357-61.
212. Gorber, S.C., et al., *The accuracy of self-reported smoking: A systematic review of the relationship between self-reported and cotinine-assessed smoking status.* Nicotine & Tobacco Research, 2009. **11**(1):12-24.
213. Willi, C., et al., *Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis.* Jama, 2007. **298**(22):2654-64.
214. Office for National Statistics. *Adult Smoking Habits in Great Britain*. 2017 19/10/2017]; Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/drugusealcoholandsmoking/datasets/adultsmokinghabitsingreatbritain.
215. National Health Service. *Quality and Outcomes Framework*. 2017 Sept 2016]; Available from: http://content.digital.nhs.uk/qof.
216. Hernán, M.A., B. Brumback, and J.M. Robins, *Marginal structural models to estimate the joint causal effect of nonrandomized treatments.* Journal of the American Statistical Association, 2001. **96**(454):440-448.
217. Mitra, R. and J.P. Reiter, *A comparison of two methods of estimating propensity scores after multiple imputation.* Statistical Methods in Medical Research, 2016. **25**(1):188-204.
218. Boggon, R., et al., *Cancer recording and mortality in the General Practice Research Database and linked cancer registries.* Pharmacoepidemiology and Drug Safety, 2013. **22**(2):168-75.
219. Pearce, N., H. Checkoway, and D. Kriebel, *Bias in occupational epidemiology studies.* Occupational and Environmental Medicine, 2007. **64**(8):562-568.
220. Sherifali, D., et al., *The effect of oral antidiabetic agents on A1C levels: A systematic review and meta-analysis.* Diabetes Care, 2010. **33**(8):1859-1864.

221. Easton, J.D., et al., *Definition and evaluation of transient ischemic attack: a scientific statement for healthcare professionals from the American Heart Association/American Stroke Association Stroke Council; Council on Cardiovascular Surgery and Anesthesia; Council on Cardiovascular Radiology and Intervention; Council on Cardiovascular Nursing; and the Interdisciplinary Council on Peripheral Vascular Disease. The American Academy of Neurology affirms the value of this statement as an educational tool for neurologists.* Stroke, 2009. **40**(6):2276-2293.
222. Farmer, R., A.S. Walker, and The Arrow Trial Team, *Marginal Structural Models for repeated measures where the intercept and slope are correlated: An application exploring the benefit of nutritional supplements on weight gain in HIV-Infected children. (Poster Presentation)*, in *UK Causal Inference Meeting*. 2016: London, UK.
223. Nichols, G.A., et al., *The change in HbA1c associated with initial adherence and subsequent change in adherence among diabetes patients newly initiating metformin therapy.* Journal of Diabetes Research, 2016. **2016**:Article ID 9687815.
224. Yusuf, S., et al., *Obesity and the risk of myocardial infarction in 27 000 participants from 52 countries: a case-control study.* The Lancet. **366**(9497):1640-1649.
225. Zhao, W., et al., *Sex differences in the risk of stroke and HbA1c among diabetic patients.* Diabetologia, 2014. **57**(5):918-926.
226. Hu, G., et al., *Body mass index, waist circumference, and waist-hip ratio on the risk of total and type-specific stroke.* Archives of Internal Medicine, 2007. **167**(13):1420-7.
227. Kurth, T., et al., *Body mass index and the risk of stroke in men.* Archives of Internal Medicine, 2002. **162**(22):2257-62.
228. Ambrose, J.A. and R.S. Barua, *The pathophysiology of cigarette smoking and cardiovascular disease: an update.* Journal of the American College of Cardiology, 2004. **43**(10):1731-7.
229. Warburton, D.E.R., C.W. Nicol, and S.S.D. Bredin, *Health benefits of physical activity: the evidence.* CMAJ : Canadian Medical Association Journal, 2006. **174**(6):801-809.
230. Suissa, S., *Immeasurable time bias in observational studies of drug effects on mortality.* American Journal of Epidemiology, 2008. **168**(3):329-35.
231. Herrett, E., et al., *Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study.* BMJ, 2013. **346**:f2350.
232. Lee, C.D., et al., *Cardiovascular events in diabetic and nondiabetic adults with or without history of myocardial infarction.* Circulation, 2004. **109**(7):855-860.
233. Roumie, C.L., et al., *Association between intensification of metformin treatment with insulin vs sulfonylureas and cardiovascular events and all-cause mortality among patients with diabetes.* JAMA, 2014. **311**(22):2288-96.
234. Yu, A.P., Y.F. Yu, and M.B. Nichol, *Estimating the effect of medication adherence on health outcomes among patients with type 2 diabetes--an application of marginal structural models.* Value Health, 2010. **13**(8):1038-45.