APPENDICES

CONTENTS

APPENDIX 1 LSHTM ETHICS APPROVAL

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636
www.lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Observational / Interventions Research Ethics Committee

Krishnan Bhaskaran
Lecturer in Statistical Epidemiology
NCDE/EPH
LSHTM

9 January 2013

Dear Dr Bhaskaran,

Study Title: **Using novel statistical methodologies to investigate the benefits and risks associated with type 2 diabetes treatments**
LSHTM ethics ref: **6349**

Thank you for your application of 4 January 2013 for the above research, which has now been considered by the Observational Committee via Chair's Action.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

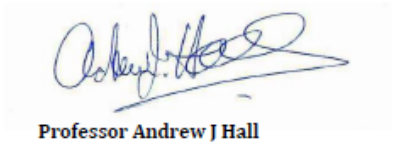
The final list of documents reviewed and approved by the Committee is as follows:

Document	Version	Date
LSHTM ethics application	n/a	04/01/2013
Diabetes cancer causal outline	6.1	04/01/2013

After ethical review

Any subsequent changes to the application must be submitted to the Committee via an E2 amendment form. All studies are also required to notify the ethics committee of any serious adverse events which occur during the project via form E4. At the end of the study, please notify the committee via form E5.

Yours sincerely,



Professor Andrew J Hall
Chair
ethics@lshtm.ac.uk
http://intra.lshtm.ac.uk/management/committees/ethics/

APPENDIX 2 ISAC APPROVAL AND PROTOCOL

Approval

ISAC EVALUATION OF PROTOCOLS FOR RESEARCH INVOLVING GPRD DATA

FEED-BACK TO APPLICANTS

CONFIDENTIAL	*by e-mail*		
PROTOCOL NO:	12_027RA		
PROTOCOL TITLE:	Using novel statistical methodologies to investigate the benefits and risks associated with type 2 diabetes treatments		
APPLICANT:	Dr Krishnan Bhaskaran, lecturer in statistical epidemiology, London School of Hygiene and Tropical Medicine, krishnan.bhaskaran@lshtm.ac.uk		
APPROVED ☒	**APPROVED WITH COMMENTS (resubmission not required)** ☐	**REVISION/ RESUBMISSION REQUESTED** ☐	**REJECTED** ☐
INSTRUCTIONS: ***Please include your response/s to the Reviewer's feedback below only if you are required to Revise/ Resubmit your protocol.*** ***Protocols with an outcome of 'Approved' or 'Approved with comments' do not require resubmission to the ISAC.*** **REVIEWER COMMENTS:** Protocol 12_027RA is approved.			
DATE OF ISAC FEEDBACK:	19/07/2016		
DATE OF APPLICANT FEEDBACK:			

For protocols approved from 01 April 2014 onwards, applicants are required to include the ISAC protocol in their journal submission with a statement in the manuscript indicating that it had been approved by the ISAC (with the reference number) and made available to the journal reviewers. If the protocol was subject to any amendments, the last amended version should be the one submitted.

Protocol

Using novel statistical methodologies to investigate the benefits and risks associated with type 2 diabetes treatments

Dr Krishnan Bhaskaran *(lecturer in statistical epidemiology[1])*
Dr Rhian Daniel (*research fellow in medical statistics[2]*)
Dr Deborah Ford (*senior statistician[3] and honorary senior lecturer[4]*)
Professor Nishi Chaturvedi *(professor of clinical epidemiology[5])*
Professor Stuart Pocock (*professor of medical statistics[2]*)
Professor Liam Smeeth (*professor of clinical epidemiology[1]*)

[1]Department of Non-Communicable Diseases Epidemiology, London School of Hygiene and Tropical Medicine, London
[2]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London
[3]MRC Clinical Trials Unit, London
[4]Department of Infection and Population Health, University College London, London
[5]National Heart and Lung Institute, Imperial College, London

Lay summary
Diabetes is often treated with drug therapies including oral anti-diabetics and insulin preparations, but quantifying the benefits and risks associated with these treatments is not straightforward. Randomised clinical trials commonly follow patients up for a relatively short time, and investigate the effects of drug therapies on markers of glucose control, rather than on clinical outcomes such as risk of death, cardiovascular disease, or cancer. Databases of routine healthcare records allow us to examine the records of much larger numbers of patients with longer follow up, but simple comparisons of patients on different treatments would give misleading results because patients given more intensive treatments are likely to be those with poorer glucose control or more advanced disease. Furthermore, because glucose control and disease severity are themselves affected by choice of treatment, standard statistical methods cannot give meaningful estimates of the underlying effects of treatment. The aim of this project is to apply a recently developed class of statistical models, known as marginal structural models, to quantify the effects of specific diabetes treatments on the risks of death, cardiovascular disease, and cancer. These models have been developed to enable treatment effects to be quantified in the presence of the problems described above.

Aims and objectives
Aim
To make use of emerging statistical methods to address pharmacoepidemiological questions regarding the benefits and risks of type 2 diabetes treatments that have been difficult to answer with conventional analyses.

Specific objectives:
1) To estimate the causal effect of specific pharmacological type 2 diabetes treatments on overall mortality, and on risk of cardiovascular events.

2) To quantify the causal effect, if any, of type 2 diabetes treatments on cancer risk.

3) To apply and extend novel statistical methods (principally, but not limited to, marginal structural models) to fulfil objectives 1 and 2 with appropriate control for time-dependent confounders that are

affected by past treatment, and to compare the results with analysis using conventional Cox regression models.

Background and rationale

Type 2 diabetes is often initially managed with non-pharmacological therapies (diet and exercise), and long-term plasma glucose levels are monitored using glycated haemoglobin (HbA1c) as a proxy. If and when HbA1c levels rise above certain threshold levels, pharmacological treatment is initiated, typically with oral metformin as the first-line therapy. Second-line and subsequent treatments can include alternative oral antidiabetics, combination therapy, and insulin.[1]

The causal effects of different treatment decisions on both beneficial outcomes (e.g. reductions in mortality and the risk of cardiovascular events) and adverse effects (e.g. increases in cancer risk) are of interest, but are not straightforward to quantify. Randomized controlled trials (RCTs) assessing the efficacy of diabetes treatments have tended to be dominated by trials with non-clinical outcomes, most commonly change in HbA1c.[2] Two trials set up in the 1970s are notable exceptions. The University Group Diabetes Program raised controversy when the results suggested an association between tolbutamide and cardiovascular mortality,[3] and later a lack of efficacy of insulin on vascular outcomes,[4] though these findings were later questioned.[5] Meanwhile, the UK Prospective Diabetes Study[6, 7] established the overall health benefits of controlling glucose status, but did not conclusively establish the relative merits of the various treatment options.[8] RCTs have also lacked the statistical power and follow-up time to detect relatively infrequent adverse events with a long lead-time, such as cancer, though meta-analysis has in some instances been used to accumulate evidence on adverse events from multiple trials, as exemplified by investigations of cardiovascular risk associated with rosiglitazone.[9]

The availability of large databases of electronic healthcare records provides an opportunity to address these research questions with much greater statistical power in an observational setting. A typical approach to assessing treatment effects in observational studies would be to include one or more time-updated treatment covariates in a Cox regression model, with adjustment for a number of baseline covariates to control for confounding. However, standard regression models produce effect estimates that have no causal interpretation when exposure is time-varying and there exists at least one covariate which is a time-dependent confounder affected by past exposure (i.e. a variable which is both predicted by prior exposure, and a predictor of subsequent outcome and subsequent exposure).[10] In assessing the causal effects of diabetes treatments on outcomes such as mortality and cancer risk, routinely collected monitoring measures such as HbA1c may act as time-dependent confounders, satisfying the above conditions: past treatment history is likely to be a predictor of current HbA1c, and current HbA1c may predict both the outcome and future treatment.

We plan to use recently developed classes of causal models aimed at dealing with time-varying confounding in order to quantify the causal effects of diabetes treatments on mortality, cardiovascular risk, and cancer risk, using data from the General Practice Research Database (GPRD), and treating HbA1c and other routine measures (blood pressure, cholesterol levels, body mass index) as time-varying confounders. Marginal structural models (MSMs) use inverse probability of treatment weights to recover a pseudopopulation which is unconfounded by measured covariates and in which model parameters have a causal interpretation under the assumption of no unmeasured confounders.[10] MSMs have been used successfully, notably in quantifying the causal effects of anti-HIV therapies,[11] and are increasing in popularity.[12] Other related methods have been suggested to account for time-

varying confounding, namely g-computation, and g-estimation of structural nested failure time models; these methods will also be explored.[13]

Study type
Hypothesis-generating

Study population
For each defined class of treatment (see "Exposure" below), a cohort will be selected comprising all patients with evidence of type 2 diabetes (provisional algorithm for identifying patients in appendix) and at least 1 year of intial follow-up in GPRD free of that treatment class (to exclude prevalent use of unknown duration). Analysis will be initially restricted to the calendar period from 2000, but the feasibility of utilising earlier data to obtain higher power and longer follow-up times will be explored, specifically with reference to the historical completeness of routinely collected diabetes monitoring data such as HbA1c.

Sample size/power
Methodology for calculating power associated with marginal structural models has not yet been developed, indeed investigation of this issue is a methodological area into which this work could be extended. For illustration of the likely power of these analyses, the following calculations refer to standard Cox analyses based looking at the effect of insulins, and of metformin, on cancer risk.

Feasibillity counts suggest that there are 455765 individuals in GPRD that have a record of type 2 diabetes, with a median follow-up of 3.3 years from the first record of diabetes (IQR 1.2 to 7.0). Assuming a baseline cancer incidence among diabetics of 15.7 cases per 1000 person years,[14] we would expect (0.0157*3.3) = 5.2% of patients to receive a cancer diagnosis during follow-up. Based on 15.6% of type 2 diabetics having a record of insulin use (GPRD feasibility count), we would have just under 80% power to detect a hazard ratio for cancer associated with insulin use of at least 1.05, and greater than 90% power to detect a hazard ratio of at least 1.06. Based on 53.2% of type 2 diabetics having a record of metformin use (GPRD feasibility count), we could detect smaller hazard ratios for this exposure (HR>1.04 [or <0.96] with 80% power, and HR >1.05 [or < 0.96] with 90% power).

Exposure, outcome, and covariates
Exposure
Benefits and risks of the following specific drugs/drug classes will be examined, initially in separate analyses: metformin, sulfonylureas, thiazolidinediones (TZDs), insulin. Exposure will be based on prescription data and specific drugs will be classified based on their place in the British National Formulary (BNF).[15]

Changes/switches of therapy
Diabetic patients may switch therapies several times during the course of their disease. We will conduct descriptive analysis to investigate the extent and nature switching. For the main analysis of each drug class, we will focus on the effect of treatment initiation; subsequent switches or interruptions of therapies will be ignored in the analysis. For example, when investigating the use of insulin, those starting insulin will be assumed to stay on insulin indefinitely. This will lead to effect estimates with an interpretation regarding the causal effect of starting treatment (analogous to intention to treat effect estimates in the context of clinical trials).

Subsequent analysis will aim to incorporate all treatments in a single model and will require diabetes treatments to be divided into broad and mutually exclusive treatment categories; the choice of categories will be guided by an initial description of treatment use in practice, but could for example include:

- Metformin monotherapy
- Sulfonylurea monotherapy
- Metformin/sulfonylurea combination therapy without a TZD or insulin
- TZD-containing regimen without insulin
- Insulin-containing regimen

Where numbers allow we will investigate the role of specific drugs within classes, but any such analyses will be presented as secondary/exploratory.

Outcome
Three specific classes of outcomes will be examined:

- All-cause mortality
- Incident cardiovascular disease
    - Stroke and myocardial infarction (fatal and non-fatal)
    - Heart failure
- Incident cancer
    - All cancer excluding non-melanoma skin cancer
    - The most common site-specific cancers (specifically breast, lung, colorectal, prostate)
    - Other cancers known/suspected to be related to diabetes (specifically pancreas, uterus, bladder, liver)

For the primary analysis, these outcomes will be ascertained directly from GPRD. Code lists for cardiovascular and cancer endpoints have been previously developed and will be updated with the involvement/review of a GP on the study team. In sensitivity analyses, the main models will then be re-run using linked data for practices that have linkages available. The linked data to be used are ONS data for mortality, MINAP and HES for cardiovascular disease, and cancer registry and HES for cancer.

Covariates
We will develop a full list of potential confounders through initial literature review. As a starting point, the following time-varying confounders will be included in the treatment prediction model, and thus accounted for in the final marginal structural model:

- HbA1c
- Cholesterol (LDL, HDL, triglycerides) (where available)
- BMI
- Blood pressure
- Serum creatinine/eGFR (where available)
- Recent anaemia
- Recent oedema
- Recent proteinuria
- Recent hypoglycaemic episode
- Recent cardiovascular event
- Calendar year

- Treatment history

(Recent defined as in the last 6 months, cut off subject to sensitivity analysis)

The following time-fixed confounders will also be included:

- Age (as the primary timescale)
- Gender
- Ethnicity (where available)
- Socioeconomic status

Linkages required
Cancer registry, MINAP, hospital episode statistics (HES), ONS mortality, individual socioeconomic status data

Data management and computing
Data management will be carried out using Stata statistical software. A high memory high spec PC (4-core, 72GB memory) is available, as is a high performance cluster (32 x 4-core nodes, capable of running 256 parallel jobs).

Statistical analysis
Descriptive analysis of treatment patterns
A descriptive analysis of diabetes treatments will be carried out to inform subsequent analyses including categorisation of treatments and development of a treatment prediction model. Person-time on specific drug classes and combinations will be tabulated univariately and by other covariates.

Marginal structural modelling
Marginal structural Cox models[10] will then be developed to estimate the causal effect of treatment on each outcome, with age as the primary timescale. Follow-up will begin at the date entry into GPRD (plus 1 year free of the treatment of interest), or if later at the first evidence of diabetes; and will end at the earliest of: occurrence of the outcome, the last data collection date for the practice, transfer out of GPRD, death. The effects of each of the major treatment classes (metformin, sulfonylureas, TZDs, insulin) will first be considered separately, with the remaining treatment classes included as additional time-varying confounders. Causal effects will be estimated based on both binary treatment (effectively comparing a policy of "always treat" with "never treat"), and cumulative treatment (estimating the causal effect associated with each additional month of treatment). A combined treatment model will then be explored, based on mutually exclusive treatment categories.

To enable model fitting with standard software, the parameters of the MSMs will be estimated by expanding each participant's record into individual person-months of follow-up and fitting a pooled logistic regression model,[16] which allows the necessary time-varying inverse probability of treatment weights (IPTW) and censoring weights to be used (see below). The model will include a cubic spline basis for age, effectively allowing the baseline hazard to take on a flexible form as in straightforward Cox regression.

Estimation of weights for the MSM
For the single treatment class models, the inverse probability of treatment weights will be estimated by fitting a logistic regression model to the person-month data, with a binary outcome variable representing "on the specified treatment in that month". The model will include the time-varying and time-fixed covariates listed under "Exposure, outcome, and covariates". Predictions from the model will be used to calculate the probability of the observed treatment history and hence the IPTWs.

Weights will be stabilised and large weights (>10) truncated as has been done in previous studies to avoid undue influence of a small number of observations.[11] Censoring weights will be estimated in a similar way. For the combined treatment model, multinomial logistic regression will be used to generated the predicted probabilities of observed treatment history and hence the treatment weights.

Other methods and comparison of models
We will compare the application and results of the above MSM analysis with two related methods for causal inference in the presence of time-varying confounding affected by past exposure, namely g-computation,[17] and g-estimation of structural nested failure time models;[13] we will also compare our results to standard Cox models to illustrate the effect of properly accounting for time-varying confounding.

Multiple testing issues
Since there are a number of exposures and outcomes, some "significant" results might be expected purely due to chance. In keeping with the hypothesis generating nature of this study, we will focus on the presentation of confidence intervals and their appropriate interpretation in the light of the number of results presented, and we will interpret p-values in terms of strength of evidence (again in the light of the number of comparisons presented), rather than employing formal p-value cut-offs.[18]

Validation of marginal structural models using known associations from randomised trial evidence
Marginal structural models are a relatively recent development in statistics, but have now been used successfully in mainstream epidemiology, notably in HIV research.[11, 19, 20] Investigators have successfully replicated clinical trial findings,[21] and explored questions that could not be addressed by trials. [11] The replication of a trial finding is a useful mechanism for developing appropriate model specifications, establishing the validity of the methodology, and increasing confidence in later results. We therefore intend to carry out a preliminary analysis replicating an established clinical trials finding in the context of diabetes therapy. There is a lack of trial data with clinical endpoints and simple comparison groups investigating the main glucose-lowering therapies (which is part of the motivation for this programme of work), but such trials do exist for other treatments used in the diabetes population. The Collaborative Atorvastatin Diabetes Study (CARDS) investigated the role of statin use for the primary prevention of cardiovascular events in type 2 diabetic patients, finding a 36% reduction in acute coronary heart disease events compared with placebo.[22] We will attempt to replicate this finding within GPRD using a marginal structural model in a similar patient population, with LDL cholesterol measures included as a time-dependent confounder.

Study limitations
We plan to adjust for a wide range of potential confounding factors. Nevertheless, it is possible that other factors influencing treatment decisions will be missed or indeed are not captured in GPRD – there could thus be residual confounding by indication. Patterns of confounding by indication may furthermore vary with time, for example as safety concerns emerge, however we will be able to explore this through including interactions with calendar time. More generally, we will rely on an untestable assumption of no unmeasured confounders. Ascertainment of outcomes based purely on GPRD data could lead to misclassification, but we will address this in sensitivity analyses in which we will attempt to confirm our results using alternative sources of outcome data, namely linked data from ONS, MINAP, HES and the cancer registries. Specific potential methodological problems include instability of weights in the MSM-fitting process, which we will address through the use of stabilised weights and weight truncation; the need for treatment groups to be balanced with respect to time-dependent confounders influencing outcome in the re-weighted pseudo-population, which we will assess descriptively. The modelling framework also require positivity (whereby patients must have a

non-zero probability of each treatment option); should this condition fail, we may need to broaden the categories of treatment groups being investigated. The study should have good generalisability to the broader UK population, as our study population will include all patients with a record of type 2 diabetes in GPRD during the study period, other than those in whom treatment commenced near or prior to GPRD registration.

Patient and user group involvement

We do not believe this research would benefit from patient group involvement at this stage, although will actively collaborate with such groups in the dissemination strategy.

Plans for dissemination

Clinical findings will be disseminated through presentation at international conferences, and through publication in the relevant medical journals. We also plan to publish our methodological findings in epidemiology or medical statistics journals.

provisional algorithm for identification of patients with type 2 diabetes

Type 2 diabetes patients will be identified by the following algorithm, based on the Royal College of General Practitioners/NHS diabetes coding report[A1]:

All patients with a code specifically indicating type 2 diabetes
+
All patients with a code indicating diabetes of unspecified type and first 6 months treatment with diet or oral agents only (i.e. no codes suggesting continuous insulin in this time)
+
All patients with a code indicating diabetes of unspecified type and aged 35+ years at first code indicating diabetes

Codelists for type 2 diabetes and diabetes of unspecified type will be based upon those developed as part of the Cardiovascular disease research using Linked Bespoke studies and Electronic Records (CALIBER) programme (http://www.caliberresearch.org/), and will be reviewed by LS and NC.

Appendix References

A1. Royal College of General Practitioners. Coding, classification, and diagnosis of diabetes. NHS Diabetes, Department of Health, 2011.

References

1. National Collaborating Centre for Chronic Conditions, *NICE clinical guideline 87 (partial update of NICE clinical guideline 66: The management of type II diabetes)*. 2009, National Institute for Healthcare and Clincal Excellence.
2. National Collaborating Centre for Chronic Conditions, Type 2 Diabetes: national clinical guideline for management in primary and secondary care (update) 2008, NICE: London.
3. Meinert, C.L., et al., *A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. II. Mortality results.* Diabetes, 1970. **19**: p. Suppl:789-830.
4. Knatterud, G.L., et al., *Effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. VII. Mortality and selected nonfatal events with insulin treatment.* Jama, 1978. **240**(1): p. 37-42.
5. Kilo, C., J.P. Miller, and J.R. Williamson, *The Achilles heel of the University Group Diabetes Program.* Jama, 1980. **243**(5): p. 450-7.
6. UK Prospective Diabetes Study (UKPDS) Group, *Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group.* Lancet, 1998. **352**(9131): p. 854-65.
7. UK Prospective Diabetes Study (UKPDS) Group, *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.* Lancet, 1998. **352**(9131): p. 837-53.
8. Nathan, D.M., *Some answers, more controversy, from UKPDS. United Kingdom Prospective Diabetes Study.* Lancet, 1998. **352**(9131): p. 832-3.
9. Nissen, S.E. and K. Wolski, *Rosiglitazone revisited: an updated meta-analysis of risk for myocardial infarction and cardiovascular mortality.* Arch Intern Med, 2010. **170**(14): p. 1191-1201.
10. Robins, J.M., M.A. Hernan, and B. Brumback, *Marginal structural models and causal inference in epidemiology.* Epidemiology, 2000. **11**(5): p. 550-60.

11. Ray, M., et al., *The effect of combined antiretroviral therapy on the overall mortality of HIV-infected individuals.* AIDS, 2010. **24**(1): p. 123-37.
12. Suarez, D., R. Borras, and X. Basagana, *Differences between marginal structural models and conventional models in their exposure effect estimates: a systematic review.* Epidemiology, 2011. **22**(4): p. 586-8.
13. Hernan, M.A., et al., *Structural accelerated failure time models for survival analysis in studies with time-varying treatments.* Pharmacoepidemiol Drug Saf, 2005. **14**(7): p. 477-91.
14. Hense, H.W., et al., *Cancer incidence in type 2 diabetes patients - first results from a feasibility study of the D2C cohort.* Diabetol Metab Syndr, 2011. **3**(1): p. 15.
15. British National Formulary, *BNF Volume 61*. 2011: BMJ Group/Pharmaceutical Press.
16. D'Agostino, R.B., et al., *Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study.* Stat Med, 1990. **9**(12): p. 1501-15.
17. Taubman, S.L., et al., *Intervening on risk factors for coronary heart disease: an application of the parametric g-formula.* Int J Epidemiol, 2009. **38**(6): p. 1599-611.
18. Sterne, J.A. and G. Davey Smith, *Sifting the evidence-what's wrong with significance tests?* Bmj, 2001. **322**(7280): p. 226-31.
19. Cain, L.E., et al., *When to initiate combined antiretroviral therapy to reduce mortality and AIDS-defining illness in HIV-infected persons in developed countries: an observational study.* Ann Intern Med, 2011. **154**(8): p. 509-15.
20. Hernan, 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): p. 561-70.
21. Delaney, J.A., S.S. Daskalopoulou, and S. Suissa, *Traditional versus marginal structural models to estimate the effectiveness of beta-blocker use on mortality after myocardial infarction.* Pharmacoepidemiol Drug Saf, 2009. **18**(1): p. 1-6.
22. Colhoun, H.M., et al., *Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial.* Lancet, 2004. **364**(9435): p. 685-96.

Amendments – 15th July 2016

1. Additional investigators:

- Ruth Farmer, PhD Candidate, Dept. Non Communicable diseases epidemiology, London School of Hygiene and Tropical Medicine

2. Aims and objectives (additional)

- An additional aim to the original study is to compare dynamic treatment strategies of the form "treat when hba1c raises above x%" (for x = 6.5%, 7%, 8%,…) in terms of impact on risk of myocardial infarction (MI), stroke, glucose control (as measured by repeated HbA1c) and overall mortality.
- This addition allows us to examine the use of dynamic marginal structural models (see amendment to statistical methods below) , an alternative application of the IPTW methodology already described, to look at research questions involving dynamic rather than static treatments for T2DM. This extends and enhances the original protocol objective of comparing static regimes only (e.g. Treat with metformin vs don't treat with metformin) on cardiovascular events and mortality.

2. Outcomes (additional)

- Hba1c will be additionally examined as a repeated measures outcome in order to study the effect of treatment on long term glucose control.
- An additional endpoint of time to target HbA1c (which will vary between 6-7% for sensitivity analysis) will be examined as an outcome for assessing the effect of different dynamic treatment regimes.

3. Statistical methods

Marginal structural modelling (change to original)

- Originally the intention was to fit the MSM's in all prevalent and incident type 2 diabetes patients. However, to reduce impact of confounding by disease severity that cannot be entirely adjusted for using the measured covariates, we now restrict the analysis to incident diabetics. This was defined as having at least 12 months of follow up in CPRD before the first diabetes related code, as defined by prescription for any diabetic agent, or a clinical code relating to diagnosis of or care for type 2 diabetes. As stated in the original amendment we follow and intention to treat principle such that once a subject starts their initial therapy it is assumed fixed. However instead of time updating to allow for subsequent changes in therapy we have censored at change to subsequent therapy and applied inverse probability of censoring weights to account for this [23] This approach eliminates concomitant diabetes medications as time dependent confounders and aids interpretation of the estimated effects as the causal effects of a specific monotherapy option.

Dynamic Marginal Structural Models (addition to original)

- Treatment decisions which are made in response to measures of disease severity that vary through time (i.e. HbA1c level) are called dynamic treatment strategies. For example, we may want to compare a range of strategies of the form "treat when HbA1c raises above x%", where x take the values 6.5%, 7%, 8% etc. Dynamic marginal structural models are a class of MSMs that in theory allow comparison of such treatment regimens with a causal interpretation. In practice, they are constructed by censoring a subject at the time they no longer follow the specified strategy, with inverse probability weights to estimate the probability of remaining uncensored in each time interval. This creates a pseudo population where all patients adhere to the specified strategy. This process can be repeated for multiple strategies and the outcomes under each strategy compared. Weights for these models will be calculated in the same way as the inverse probability of treatment weights as described in the original protocol. The use of grace periods will also be investigated, this allows patients to "initiate within m months of first HbA1c measure above x%" and may increase adherence to regimes and therefore reduce the numbers censored[24]. Different values of m will be investigated to see to what extent allowing such grace periods affects a) adherence and b) parameter estimates.

Validation of marginal structural models using known associations from randomised trial evidence (change to original)

- Instead of replicating the CARDS trial as stated in the protocol, we decided after discussion among the investigators to focus on attempting to replicate the findings of the UKPDS study[6], using MSM with IPTW to attempt to recreate the results showing superiority of metformin and sulfonylurea to diet in terms of risk of MI, and stroke, and also in terms of predicted long term HbA1c levels. The reason for this change was that it was decided that this would fit in better with the overall focus of our work (and the PhD of which it is a part) on diabetes-specific treatments.

Additional References:

23. 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): p. 779-88.
24. Cain, L.E., et al., *When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data.* Int J Biostat, 2010. **6**(2): p. Article 18.

APPENDIX 3 PUBLISHED VERSION OF SYSTEMATIC REVIEW (CHAPTER 2)

International Journal of Epidemiology, 2017, 728–744
doi: 10.1093/ije/dyw275
Advance Access Publication Date: 12 December 2016
Original article

Miscellaneous

Metformin and cancer in type 2 diabetes: a systematic review and comprehensive bias evaluation

Ruth E Farmer,[1,2]* Deborah Ford,[2] Harriet J Forbes,[1] Nishi Chaturvedi,[3] Richard Kaplan,[2] Liam Smeeth[1] and Krishnan Bhaskaran[1]

[1]Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK, [2]MRC Clinical Trials Unit at UCL, London, UK and [3]Institute of Cardiovascular Science, University College London, London, UK

*Corresponding author. Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: Ruth.farmer@lshtm.ac.uk

Accepted 2 September 2016

Abstract

Background: Existing observational studies provide conflicting evidence for the causal effect of metformin use on cancer risk in patients with type-2 diabetes, and there are concerns about bias affecting a number of studies.

Methods: MEDLINE was used to identify observational studies investigating the association between metformin and overall or site-specific cancer in people with type-2 diabetes. A systematic data extraction and bias assessment was conducted, in which risk of eight bias domains (outcome, exposure, control selection, baseline confounding, time-dependent confounding, immortal time, missing data, censoring methods) were assessed against pre-defined criteria, and rated as unlikely, low, medium or high.

Results: Of 46 studies identified, 21 assessed the effect of metformin on all cancer. Reported relative risks ranged from 0.23 to 1.22, with 12/21 reporting a statistically significant protective effect and none a harmful effect. The range of estimates was similar for site-specific cancers; 3/46 studies were rated as low or unlikely risk of bias in all domains. Two of these had results consistent with no effect of metformin; one observed a moderate protective effect overall, but presented further analyses that the authors concluded were inconsistent with causality. However, 28/46 studies were at risk from bias through exposure definition, 22 through insufficient baseline adjustment and 35 from possible time-dependent confounding.

Conclusions: Observational studies on metformin and cancer varied in design, and the majority were at risk of a range of biases. The studies least likely to be affected by bias did not support a causal effect of metformin on cancer risk.

Key words: Pharmacoepidemiology, diabetes, cancer, confounding, bias, causality

Key Messages

- Many existing observational studies investigating the effect of metformin use on cancer incidence in patients with type 2 diabetes have risk of bias.
- No studies to date have used appropriate statistical models to estimate the effect of time-varying treatment correctly controlling for time-dependent confounders which may be affected by previous treatment.
- Studies at lowest risk of bias do not support the hypothesis that metformin is protective against cancer.
- Previously reported large protective associations are unlikely to be causal.

Introduction

Research exists to suggest type 2 diabetes mellitus (T2DM) may be a risk factor for cancer,[1,2] and observational studies have suggested that diabetic therapies could also affect this risk.[3–5] Multiple observational studies have reported an apparent protective effect of metformin, a common first-line therapy for T2DM, against incidence of any cancer.[5–9] However, a number of potential biases have been identified within some of these studies.[10] There are limited data from clinical trials comparing metformin with other treatments, though one meta-analysis of adverse events from trials did not find any association between metformin use and cancer occurrence.[11].

Particular difficulties arise for observational studies in this context because treatment with metformin for T2DM changes through time (is 'time varying'), and is influenced by disease severity. This means that disease severity may be a confounder between metformin use and cancer, but will also be on the causal pathway since metformin is prescribed in order to control disease severity. For example, glycated haemoglobin (HbA1c), a measure of long-term blood glucose control, and body mass index (BMI) are predictive of metformin use according to well-defined treatment guidelines for T2DM,[12] but use of metformin will likely influence future HbA1c and BMI. There is also evidence that both BMI[13] and HbA1c[14] affect cancer risk. In this situation, standard statistical models cannot estimate the true causal effect of time-varying treatment.[15] Throughout this paper, such time-updated variables that may be both confounders of, and on the causal pathway for the association between exposure and outcome are referred to as 'time dependent confounders affected by prior treatment' (Box 1).

Reviews to date have examined existing evidence for the link between metformin use and cancer; however, some were not comprehensive[10] and others have not systematically evaluated or presented a detailed evaluation of bias.[16–21]

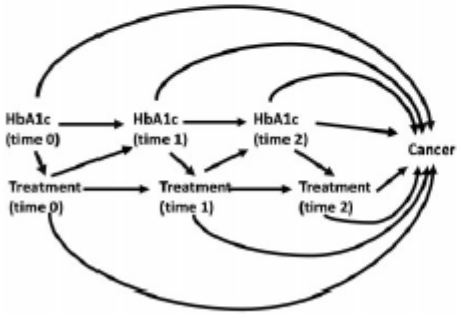
The aim of this study was to summarize existing observational studies investigating possible associations between metformin use and cancer risk in patients with T2DM, and to systematically examine the research design and analysis methods with regard to risk of bias. A secondary aim was to use meta-regression to estimate the extent to which these biases may account for the differences between study estimates.

BOX 1. Key definition: 'time-dependent confounder affected by previous treatment

A variable is a time-dependent confounder if it satisfies the following conditions:

i. the variable changes through time;
ii. values are predictive of treatment initiation;
iii. the variable is also associated with the outcome of interest.

When the time-dependent confounder is also affected by previous treatment, as depicted in the causal diagram below, standard statistical methods cannot provide unbiased estimates of the total causal effect of time-varying treatment.



Methods

Search strategy

MEDLINE was searched by R.F. using OvidSP on 30 May 2014 for all English-language published articles on cancer

risk and type 2 diabetes treatments from 1946 onwards. The search involved using MeSH headings as well as keyword searches in the title and abstract. The full search terms are presented in Supplementary data (available at *IJE* online). Conference abstracts and unpublished studies were excluded.

Screening strategy

Articles were included in the review if they were of a standard epidemiological design and presented original observational research. Reviews and meta-analysis were not included. Studies were required to present a measure of 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 comorbidities or diseases were excluded.

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, for example 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 screening was then applied to the remaining papers, and the reference lists of relevant reviews and meta analyses searched. A 10% random sample of the extracted studies were screened by an additional researcher (H.F.) to test the reliability of the inclusion criteria. A Cohens kappa score[22] was calculated to give a quantitative measure of rater reliability, with a value of 0.75 used as the threshold for 'excellent agreement'.[23].

Data extraction and bias assessment

The data extraction table was piloted on five studies (by R.F., K.B. and D.F.) and subsequently refined to ensure systematic documentation of the relevant information. An example extraction table is supplied in Supplementary Table 1 (available as Supplementary data at *IJE* online).

Although none of the investigators were blinded to the aims of the review, detailed criteria for assessment of bias were produced in order to consider risks of bias for each study. The eight domains assessed for bias were: (i) outcome definition; (ii) exposure definition (including choice of comparator): (iii) control selection (case control studies only); (iv) consideration of HbA1c, BMI and other antidiabetic drugs as time-dependent confounders affected by previous treatment (Box 1); (v) adjustment for baseline (study entry) confounders (smoking, diabetes severity, age, gender); (vi) immortal time (cohort studies only); (vii) missing data; and (viii) censoring methods (cohort studies only). For each bias domain, pre-defined criteria allowed categorization into high, medium, low or unlikely risk of bias. Bias in study estimate occurs when aspects of the design or data analysis either induce or fail to eliminate non-causal imbalances in risk of cancer between those who are exposed or unexposed. How this may occur is dependent on the bias domain in question, and detailed criteria for each domain are presented in Supplementary Table 2 (available as Supplementary data at *IJE* online). Broadly, studies were considered at 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 potential magnitude of the bias was unlikely to materially affect the overall study conclusions. Medium and high risk of bias meant that there was potential for some or substantial effect of bias on the study estimate, respectively. Although the specific criteria for each bias domain may have left some room for subjectivity, they were developed and agreed in advance by R.F., K.B. and D.F. to make them as objective as possible.

Time-dependent confounders affected by previous treatment were considered as a separate domain in addition to baseline confounding, to highlight the difference between baseline confounding that could be easily adjusted for in a standard analysis, and the more subtle bias that may arise if time-dependent confounders affected by previous treatment are not correctly adjusted for. 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 whether the definitions may introduce selection bias. 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 (five 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 (five studies), a middle category best representing a moderate level of exposure was taken.

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 (three 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, other oral antidiabetic drugs (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 variable was used to reduce sparsity. Backwards stepwise selection was used to identify which 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. Analysis was conducted using STATA v14.

Results

Search and screening

The numbers of studies included/excluded at each stage of the process are presented in Figure 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 Cohen's 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 randomized controlled trials and 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 randomized.

Table 1 summarizes the data sources, outcomes, exposure definitions, timing of exposure measurements and comparator exposures used. More detailed study-level information is presented in Supplementary Table 3 (available as Supplementary data at *IJE* online).

Study characteristics

Of the 46 studies, 22 were case-control design[7,22–24] and 24 were cohort studies.[3,5,8,9,45–63] Data from electronic health records were used by 37 (80%) of the studies: most commonly, the UK's Clinical Practice Research Datalink (CPRD) (13 studies) and the Taiwan National Health Insurance Claims Database (eight studies). As previously mentioned, one paper[51] used data from two randomized controlled trials. The remaining eight (all case-control) collected data from a specific cancer or diabetes clinic.

A total of 22 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 two 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 sulphonylurea [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. A total of 21 studies examined the outcome of all cancers excluding non-melanoma skin cancer (NMSC). 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.

Effect of metformin on cancer risk

Figure 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 3.

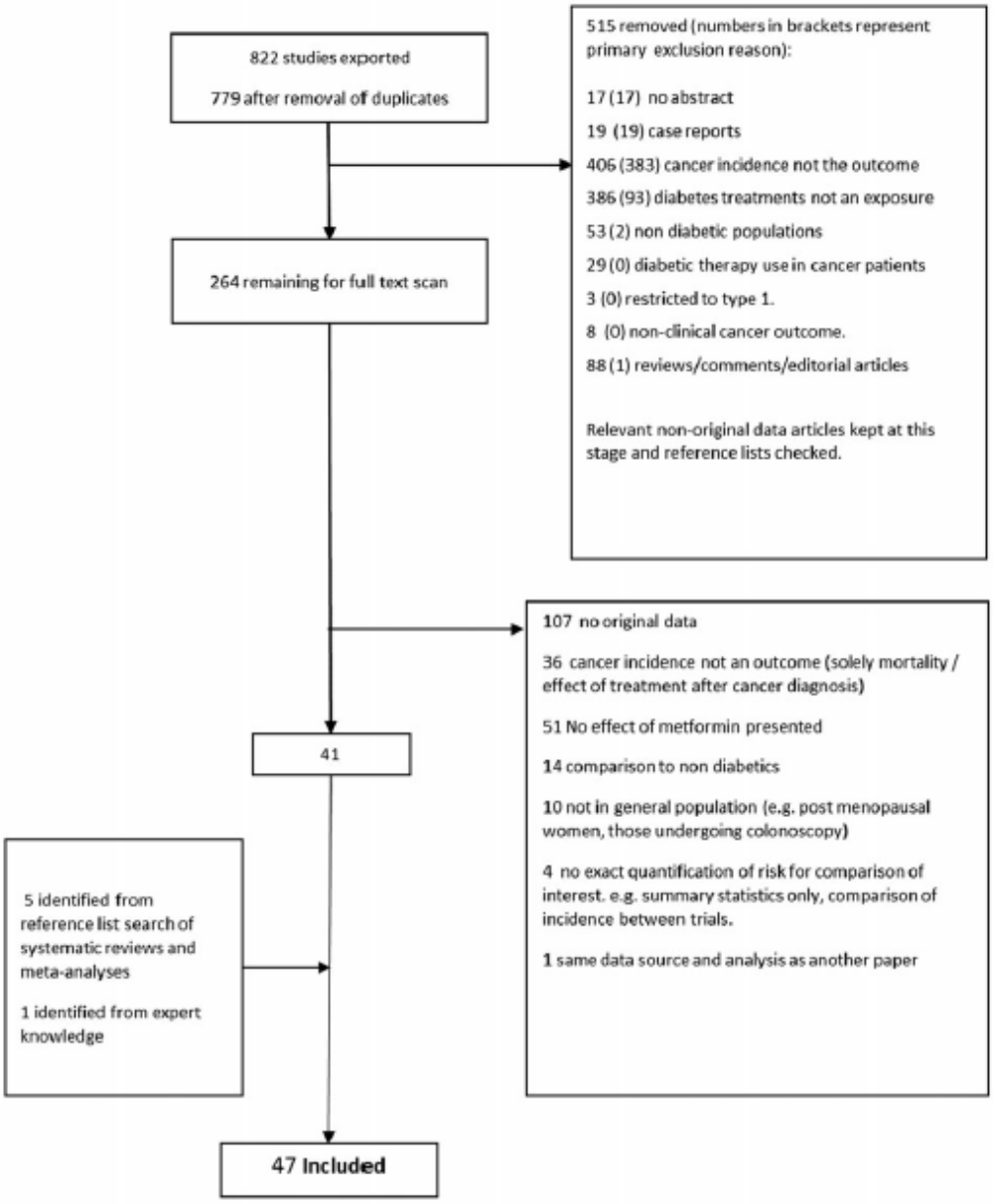


Figure 1. Flow chart of screening process detailing number of studies excluded at each stage and reason for exclusions.

For all cancer, 18/21 studies estimated a protective effect of metformin, with 12/16 having upper confidence limits below 1. The magnitude of the effect estimates ranged from just a 0.04% reduction in risk[45] to a 77% reduction in risk.[49] For site-specific cancers, estimates were also highly variable across studies (Figure 3).

Bias evaluation

Only three studies[51,61,62] 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[43] where it was rated unknown. Three of these studies saw no evidence of an effect of

Table 1. Frequency tables to summarize data source, outcome and exposure definitions for 46 studies

	Case-control *N* (%)	Cohort *N* (%)	Total *N* (%)
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[a]			
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)
Randomization	0 (0)	1 (4)	1 (2)
Combination therapy with sulphonylurea	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 ((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)
Sulphonylurea	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)
[b]New users of OADs			
Yes	3 (14)	7 (29)	10 (22)
No	17 (77)	12 (50)	29 (63)
Unsure	2 (9)	5 (21)	7 (15)

CPRD, Clinical Practice Research Datalink; GPRD, General Practice Research Database; HCC, Heaptocellular Carcinoma; ICC Intrahepatic Cholangiocarcinoma ; OAD, oral diabetic agent.

[a]Studies may have multiple outcomes; therefore column percentages will not sum to 1.

[b]Based on whether clear description is given in methods.

metformin. One study estimated a modest protective effect of long-term use (> 60 months) in comparison with short-term use (0-6 months) with a hazard ration (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.[62]

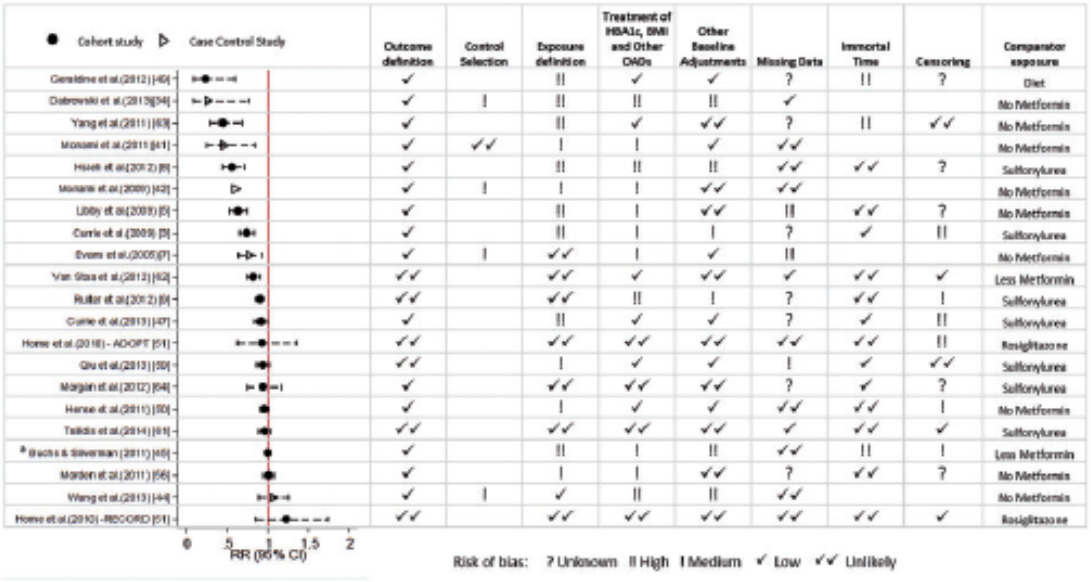


Figure 2. Estimated relative risk (odds ratio or hazard ratio) with 95% CI for the 21 studies examining use of metformin and risk of all cancers, and corresponding assessment of bias according to pre-specified criteria.
*Represents the hazard ratio for cancer risk per one extra prescription of metformin.

Study-specific results of bias assessment for studies assessing all cancer as an outcome are displayed alongside risk estimates in Figure 2. 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. Bias assessments for all other studies are presented in Supplementary Table 4 (available as Supplementary data at *IJE* online).

Time-dependent confounders affected by previous treatment (HbA1c, BMI other antidiabetics)

Only four studies were considered as unlikely to be affected by bias due to how HbA1c, BMI and other antidiabetic 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 before baseline.

Only 16/47 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[3,45] or because it did not alter the results of the multivariable model.[25,27,59,61] All but one of these studies[61] were still considered at risk since it was questionable whether the HbA1c used was representative of HbA1c at the time of starting treatment; 26 studies accounted for BMI in their final model. In most case-control studies, the measurement of HbA1c and BMI preceded 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, and 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 antidiabetic drugs is dependent upon the exposure and comparator group definitions. In six of the cohort studies examined, adjustment for use of other diabetic drugs was not necessary.[8,9,51,59,61,64] In the remaining studies, 22 accounted for OADs. Tables 2 (case-control) and 3 (cohort) detail which adjustments were made, and the timing of the measurement within the follow-up period for each study separately.

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. The exposure definition was most likely to have introduced bias in case-control studies by having different time windows to measure exposure, meaning the

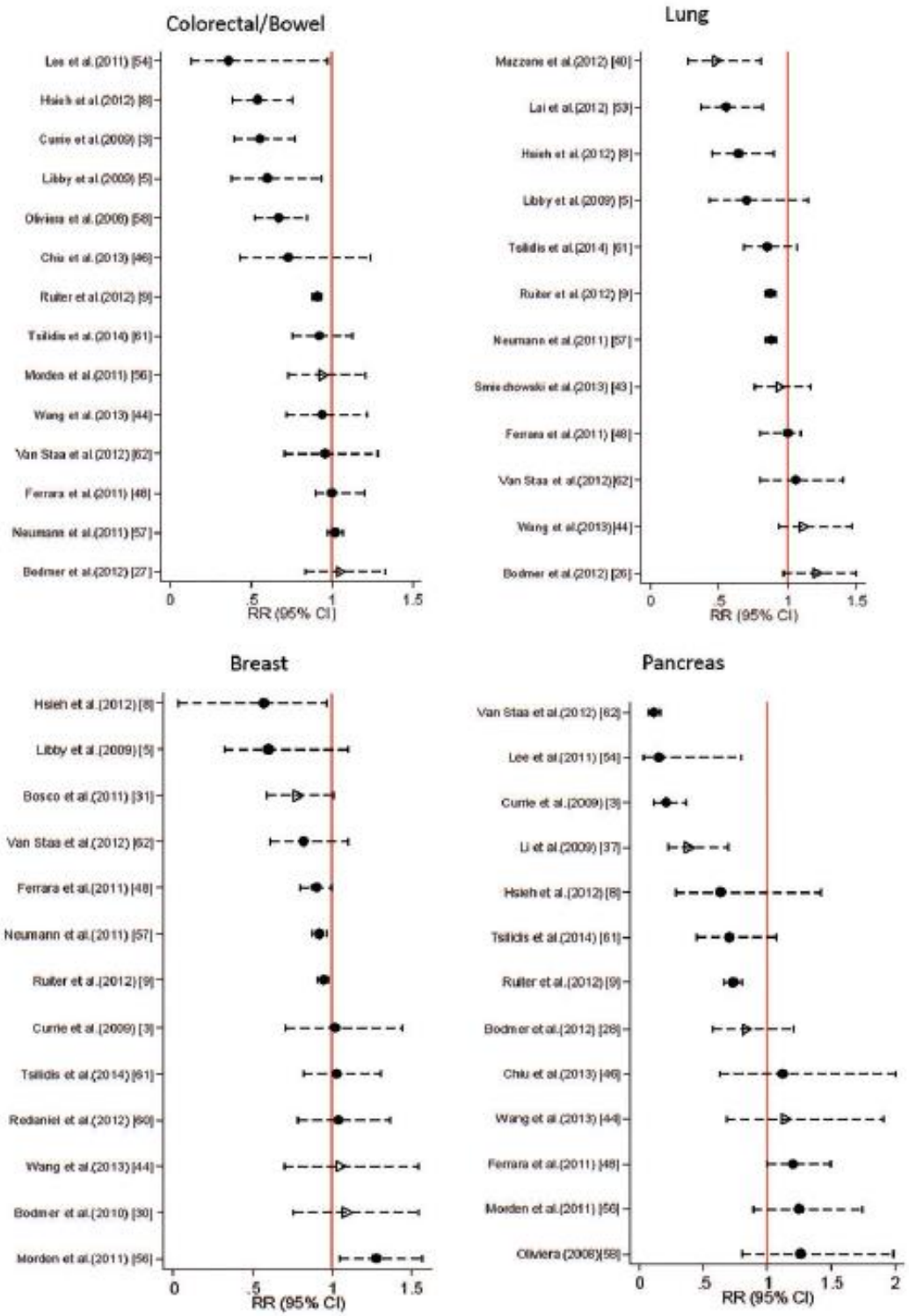


Figure 3. Estimated relative risk (odds ratio or hazard ratio) with 95% CI for 4 most commonly studied site specific cancers. Case control studies are represented by hollow triangle, Cohort studies by filled circles.

overall chance of seeing individuals exposed to metformin is systematically different between the cases and controls. 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. In all, 22 studies were considered at risk of bias from confounding due to incomplete or inappropriate baseline adjustment because either the comparator used may have resulted in comparing patients at differing disease stages without adjustment for baseline disease severity, or 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 that may have 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 was considered to be low risk. This was supported by the five studies that considered different latency periods in sensitivity analyses, concluding that estimates did not differ substantially.[9,24,43,60,61]

Many studies were considered as 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.[65] With these three studies having > 20% missing data, the effect of residual confounding could be large.

Meta-regression

Table 4 presents estimates and model diagnostics for the final meta-regression models obtained. For the outcome of all cancer, after backwards stepwise selection, the study

Table 2. Adjustment method for key time-dependent confounders affected by previous treatment: case-control studies

Study name	HbA1c			BMI			Other diabetic medication		
	Adjusted for value before first exposure	Adjusted for value between exposure and index date[a]	Adjusted for value at index date[a]	Adjusted for value before first exposure	Adjusted for value between exposure and index date[a]	Adjusted for value at index date[a]	Adjusted for value before first exposure	Adjusted for value between exposure and index date[a]	Adjusted for value at index date[a]
Azoulay *et al.* (2011)[24]		✓			✓			✓	
Becker *et al.* (2013)[25]					✓			✓	
Bodmer *et al.* (2011)[29]		✓			✓			✓	
Bodmer *et al.* (2010)[30]		✓			✓			✓	
Bodmer *et al.* (2012) (Lung)[26]					✓			✓	
Bodmer *et al.* (2012) (Pancreatic)[28]					✓			✓	
Bodmer *et al.* (2012) (Colorectal)[27]					✓			✓	
Bosco *et al.* (2011)[31]									
Chaiteerakij *et al.* (2013)[32]									
Dabrowski *et al.* (2013)[34]								✓	
Donadon *et al.* (2010)[35]			✓			✓			
Li *et al.* (2009)[37]					✓			✓	
Evans *et al.* (2005)[7]					✓				
Hassan *et al.* (2012)[38]									
Margel *et al.* (2013)[39]								✓	
Mazzone *et al.* (2012)[40]		✓			✓				
Monami *et al.* (2009)[42]			✓			✓		✓	
Monami *et al.* (2011)[41]				✓				✓	
Smiechowski *et al.* (2013)[43]	✓[b]	✓		✓[b]	✓		✓[b]	✓	
Wang *et al.* (2013)[44]									
Chen *et al.* (2013)[33]								✓	
Donadon *et al.* (2010)[36]						✓			

[a]Index date, time of cancer diagnosis/matched date for control.

[b]Sensitivity analysis assessed whether there was a difference between adjusting for covariates measured before exposure or any time between 1 year before exposure and index date.

Table 3. Adjustment method for key time-dependent confounders affected by previous treatment: cohort studies

Study name	HbA1c			BMI			Other diabetic medication		
	Adjusted for value at cohort entry (at time of or prior to first exposure)	Adjusted as a time-updated variable	Measured as an average of values at any point after exposure	Adjusted for value at cohort entry (at time of or prior to first exposure)	Adjusted as a time-updated variable	Measured as an average of values at any point after exposure	Adjusted for value at cohort entry (at time of or prior to first exposure)	Adjusted as a time-updated variable	Measured as an average of values at any point after exposure
Currie *et al.* (2009)[3]									
Currie *et al.* (2013)[47]	✓			✓					
Geraldine *et al.* (2012)[49]	✓			✓[a]					
[d]Home *et al.* (2010)[51]									
Hsieh *et al.* (2012)[8]									
Lai *et al.* (2012) (HCC)[52]									
Lai *et al.* (2012) (LUNG)[53]									
Lee *et al.* (2011)[54]									✓
Libby *et al.* (2009)[5]			✓			✓	✓[b]		
[e]Qiu *et al.* (2013)[59]									
Redaniel *et al.* (2012)[f][60]			✓	✓					
[e]Ruiter *et al.* (2012)[9]									
[e]Tsilidis *et al.* (2014)[61]				✓					
Yang *et al.* (2011)[63]	✓			✓					✓
Buchs and Silverman (2011)[45]									✓
Oliviera *et al.* (2008)[58]									
Hense *et al.* (2011)[50]				✓			✓		
Chiu *et al.* (2013)[46]									
Ferrara *et al.* (2011)[48]	✓							✓	
Lehman et al. (2012)55			✓						
Morden *et al.* (2011)[56]	✓[c]			✓[c]					
Neumann *et al.* (2011)[57]								✓	
Van Staa *et al.* (2012)[62]				✓				✓	
[e]Morgan *et al.* (2012)[64]	✓			✓					

[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.
[d]Treatment randomised so no adjustment necessary.
[e]Adjustment for use of other OADs not necessary as study looked at incident users of diabetes medications and censored at change in medication.

level predictors that remained in the model were comparator group and 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. However, this model was estimated to still have 85% residual variation due to heterogeneity. The comparator group was also retained in the models for site-specific cancers; however in the models for colorectal, lung and breast cancers, using other OADs as the comparator was estimated to make metformin appear more protective.

The strongest predictor of heterogeneity for studies of lung cancer was risk of bias from exposure definition which, if present, was estimated to reduce the log risk ratio by 0.44, 95% CI (0.17, 0.72) $P = 0.007$, making metformin appear more protective. 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 was also estimated to influence study heterogeneity for breast cancer, with presence of these biases estimated to have equal and opposite effects on the log risk ratio. For colorectal cancer, the strongest predictor was biased exposure definition, which was estimated to make metformin appear more protective.

Table 4. Parameter estimates from meta-regression models after backwards stepwise selection

		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)	[d]0.217								
	Less metformin	0.14 (−0.28, 0.55)		−0.04 (−0.43 , 0.34)	[d]0.386	0.05 (−0.34, 0.44)	[d]0.107	−0.37 (−0.81, 0.07)	[d]0.625	−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)	
Bias from time-dependent confounding	Low risk			0 (ref)				0 (ref)		0 (ref)	0.046
	High risk			0.11 (−0.06, −0.28)	0.171			0.22 (0, 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, 0.5)	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
[a]I squared		85.21%		0%		0%		3.5%		9.5%	
[b]Adjusted R^2		−20.32%		100.00%		100%		−190%		99.97%	
[c]Tau^2		0.046		0		0		0.000305		0.000156	

Estimate represents the expected change in the log risk ratio (either HR or OR, depending on analysis method) for the effect of metformin on cancer, for each study level predictor. For example, a study of metformin and lung cancer, in which there is high risk of bias from exposure definition, is estimated to have a log risk ratio 0.44 lower than a study not at risk of bias from exposure definition.

[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 p value for test of all levels comparator group.

Discussion

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; however this study included many analyses, and also reported that when comparing metformin exposure with 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 3 months after therapy initiation, which they suggested might be due to detection bias, which would also explain why longer exposure appears protective when compared with the first 6 months of therapy.

The estimates of effect reported across the 46 studies were highly variable for all outcomes studied. Many studies were at high risk of bias from exposure definition which, for reasons already outlined by Suissa and Azoulay,[10] 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. Figure 4 displays the study estimates from Figure 2 ordered by risk of bias from exposure definition (left) to demonstrate this. It is possible that confounding by disease severity, and in particular confounding from time-dependent variables affected by previous treatment, could partly explain the remaining heterogeneity in observed estimates. 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 ordered by this score [Figure 4 (right)] it is clear that heterogeneity increases as risk of bias increases, and the strongest protective effects are from those studies with the highest risk of bias overall.

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 eight domains examined. However, as in all studies of this kind, it was not possible to eliminate all subjectivity from this process.

Figure 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 previous treatment (as previously defined in Box 1). Figure 5 B, C and D illustrates the causal pathways that are actually being estimated under the three approaches most commonly used in the studies examined in this review. In Figure 5B, studies adjust for HbA1c but the measurement is taken any time during follow-up, which may result in 'adjusting out' any effect of metformin that is mediated through HbA1c. In Figure 5C, because treatment may change after baseline, the single adjustment at time 0 may lead to residual confounding by post-baseline HbA1c.

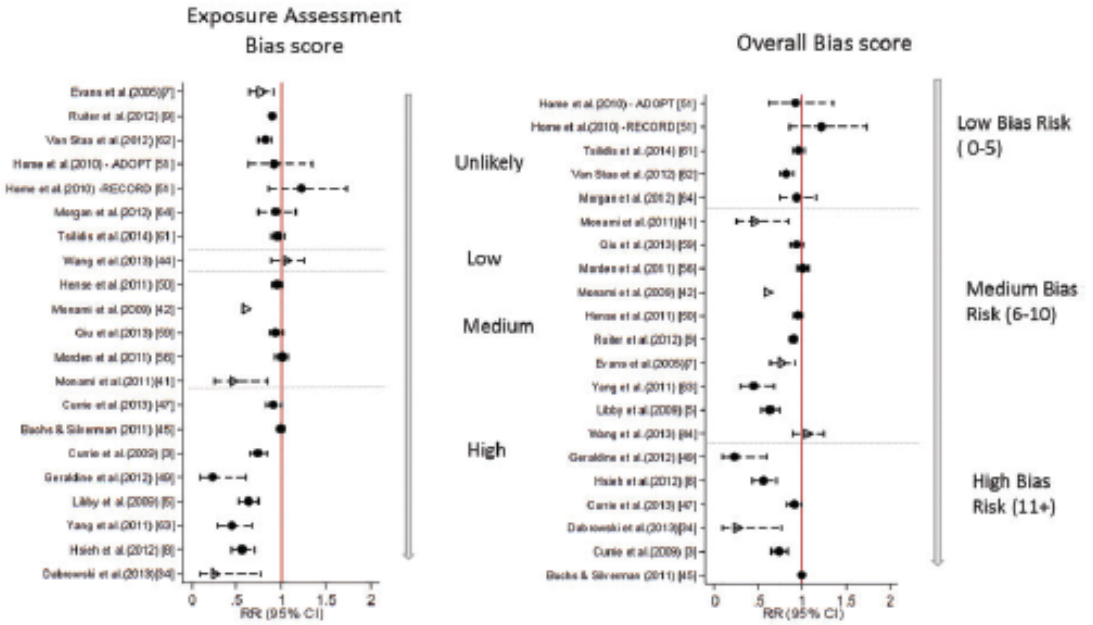


Figure 4. Estimates of relative risk of cancer from metformin use, ordered by risk of bias from exposure assessment only (left) and by overall risk of bias (right). Overall bias score is sum of bias risk over all domains, with unlikely = 0, low = 1, medium = 2, high = 3. Case control studies are represented by hollow triangle, Cohort studies by filled circle.

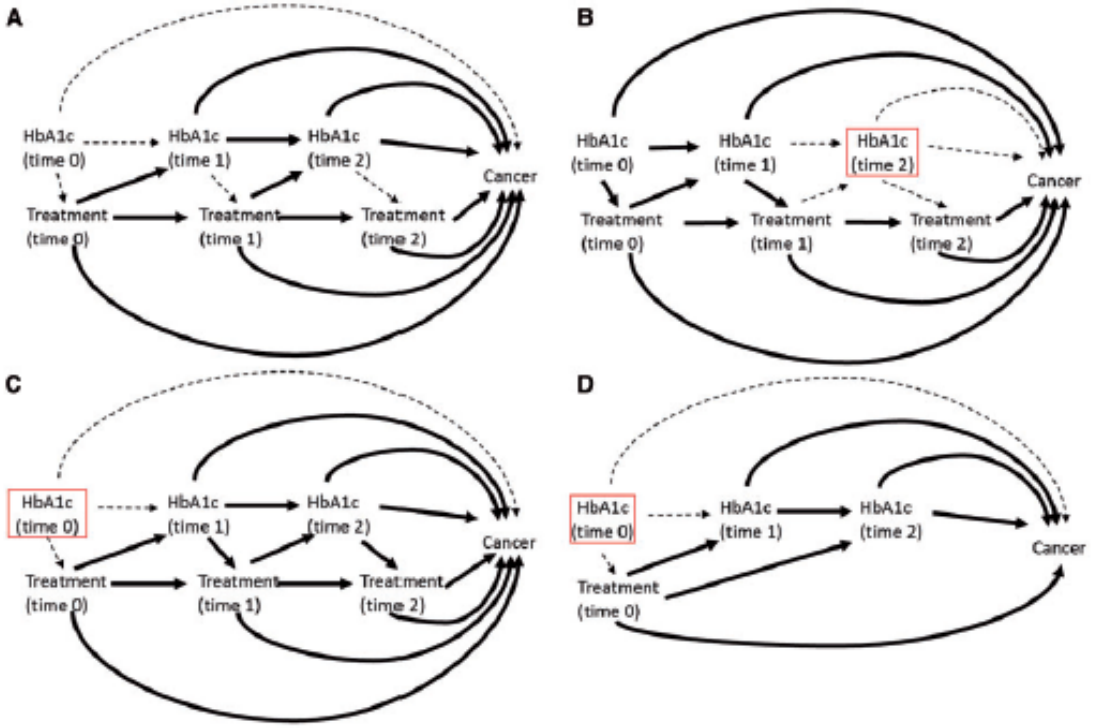


Figure 5. Directed Acyclic Graphs (DAGs) to represent estimated causal pathways for A) the desired total causal effect of treatment on cancer risk, and B)-D) the estimated effect under different methods of adjustment for time dependent confounders affected by prior treatment. Box indicates adjustment. 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.

In Figure 5D, 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, which may not be appropriate given that patients are unlikely to adhere to a single treatment throughout follow-up. One study adjusted for non-adherence[61] using a method that produces an unbiased estimate if there are no unmeasured confounders of the association between non-adherence and outcome,[66] but the validity of this assumption is questionable. This approach is also limited by considering comparisons between active drugs only. When applied and analysed carefully, it will give an unbiased estimate of the effect of initiating metformin compared with initiating (as an example) sulphonylurea 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 outlines in 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 previous treatment on the estimated effect of metformin on cancer risk.

In order to estimate the causal pharmacological effect of metformin on risk of cancer, the ideal would be to emulate a randomized controlled trial where patients are randomized 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 previous treatment on future disease severity measures. Causal methodology has been successfully used in other areas to overcome issues with

time-dependent confounders affected by previous treatment,[67,68] and could be applied to this question as a valuable addition to the current literature.

These causal inference methods (marginal structural models with inverse probability of treatment weighting or the g-computation formula) may be required to fully guard against some of the potential biases we identified, notably time-dependent confounders affected by previous treatment.[69] However, even with standard analytical approaches, careful study design and analysis can minimize the risk of bias being introduced. For example, it is desirable to clearly identify incident users of oral diabetes medications, ideally in patients with newly diagnosed diabetes, and to ensure that important confounders such as HbA1c, BMI and other disease severity measures are recorded and adjusted for at study entry–either before or at the time of medication initiation. This will ensure that disease severity is broadly balanced at study entry, and that the effect of medication on future values of important covariates is not eliminated. In addition, if medication use is not assumed to be fixed from baseline, then it is important to classify time before first exposure as unexposed in order to avoid introducing immortal time bias. Secondary analyses to look at effects of cumulative exposure, and sensitivity analysis with exclusion of periods in which un-diagnosed cancer may be affecting probability of treatment, would also be advisable to establish whether observed associations are likely to be causal.

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. Only one database was used in the search, and therefore some relevant literature may have been omitted from the review. However, by searching reference lists of other meta-analyses and systematic reviews, the majority of studies will still have been identified. Since performing the original search, it is likely that new studies will also have been published on the topic; however, a brief updated search did not identify any new studies that used methods substantially different from those covered in this review, though one study used slightly more sophisticated methods to deal with baseline confounding by indication.[70] The meta-regression aimed to establish whether any of the potential sources of bias could explain the heterogeneity in risk estimates. A comparator group was selected for the final model in all analyses as a predictor of heterogeneity, but the direction of effect was inconsistent between models. Use of a non-incident user cohort was also identified in three models as a predictor of heterogeneity, but estimates of how this would affect study results were imprecise.

The overall reliability of the meta-regression results 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 sizes 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. Also as previously mentioned, the bias evaluation could not be perfectly objective, which adds further uncertainty to any results of this analysis. Therefore, overall the results of this exploratory analysis should be interpreted cautiously.

Overall, 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 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 previous 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 is defined. Studies without such biases tend to have estimates closer to the null, and whereas 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.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

This work was supported by the Medical Research Council London Hub for Trials Methodology Research [PhD studentship to R.F.], Wellcome Trust/Royal Society Sir Henry Dale Fellowship [107731/Z/15/Z to K.B.] and Wellcome Trust Senior Research Fellowship in Clinical Science [098504/Z/12/Z to L.S.].

Conflict of interest: There is no conflict of interest to report.

References

1. Vigneri P, Frasca F, Sciacca L, Pandini G, Vigneri R. Diabetes and cancer. *Endocr Relat Cancer* 2009;16:1103-23.
2. Giovannucci E, Harlan DM, Archer MC *et al*. Diabetes and cancer: a consensus report. *CA Cancer J Clin*. 2010;60:207-21.
3. Currie CJ, Poole CD, Gale EA. The influence of glucose-lowering therapies on cancer risk in type 2 diabetes. *Diabetologia* 2009;52:1766-77.
4. Hemkens LG, Grouven U, Bender R *et al*. Risk of malignancies in patients with diabetes treated with human insulin or insulin analogues: a cohort study. *Diabetologia* 2009;52:1732-44.
5. Libby G, Donnelly LA, Donnan PT, Alessi DR, Morris AD, Evans JM. New users of metformin are at low risk of incident cancer: a cohort study among people with type 2 diabetes. *Diabetes Care* 2009;32:1620-25.
6. Baur DM, Klotsche J, Hamnvik OP *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. *Metab Clin Exp* 2011;60:1363-71.
7. Evans JM, Donnelly LA, Emslie-Smith AM, Alessi DR, Morris AD. Metformin and reduced risk of cancer in diabetic patients. *BMJ* 2005;330:1304-05.
8. Hsieh MC, Lee TC, Cheng SM, Tu ST, Yen MH, Tseng CH. The influence of type 2 diabetes and glucose-lowering therapies on cancer risk in the Taiwanese. *Exp Diabetesity Res* 2012;2012:413782.
9. Ruiter R, Visser LE, van Herk-Sukel MP *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;35:119-24.
10. Suissa S, Azoulay L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care* 2012;35:2665-73.
11. Stevens RJ, Ali R, Bankhead CR *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;55:2593-603.
12. National Collaborating Centre for Chronic Conditions. *Type 2 Diabetes: National Clinical Guideline for Management in Primary and Secondary Care (Update)*. London: NICE, 2008.
13. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. *Lancet* 2014;384:755-65.
14. de Beer JC, Liebenberg L. Does cancer risk increase with HbA1c, independent of diabetes? *Br J Cancer* 2014;110:2361-68.
15. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11:550-60.
16. Franciosi M, Lucisano G, Lapice E, Strippoli GF, Pellegrini F, Nicolucci A. Metformin therapy and risk of cancer in patients with type 2 diabetes: systematic review. *PLoS One* 2013;8:e71583.
17. Decensi A, Puntoni M, Goodwin P *et al*. Metformin and cancer risk in diabetic patients: a systematic review and meta-analysis. *Cancer Prev Res (Phila)* 2010;3:1451-61.
18. Gandini S, Puntoni M, Heckman-Stoddard BM *et al*. Metformin and cancer risk and mortality: a systematic review and meta-analysis taking into account biases and confounders. *Cancer Prev Res (Phila)* 2014;7:867-85.
19. Noto H, Goto A, Tsujimoto T, Noda M. Cancer risk in diabetic patients treated with metformin: a systematic review and meta-analysis. *PLoS One* 2012;7:e33411.
20. Wang Z, Lai ST, Xie L *et al*. Metformin is associated with reduced risk of pancreatic cancer in patients with type 2 diabetes mellitus: a systematic review and meta-analysis. *Diabetes Res Clin Pract* 2014;106:19-26.
21. Zhang ZJ, Zheng ZJ, Shi R, Su Q, Jiang Q, Kip KE. Metformin for liver cancer prevention in patients with type 2 diabetes: a systematic review and meta-analysis. *J Clin Endocrinol Metab* 2012;97:2347-53.
22. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37-46.
23. Banerjee M, Capozzoli M, McSweeney L, Sinha D. Beyond kappa: A review of interrater agreement measures. *Can J Stat* 1999;27:3-23.
24. Azoulay L, Dell'Aniello S, Gagnon B, Pollak M, Suissa S. Metformin and the incidence of prostate cancer in patients with type 2 diabetes. *Cancer Epidemiol Biomarkers Prev* 2011;20:337-44.
25. Becker C, Jick SS, Meier CR, Bodmer M. Metformin and the risk of endometrial cancer: a case-control analysis. *Gynecol Oncol* 2013;129:565-69.
26. Bodmer M, Becker C, Jick SS, Meier CR. Metformin does not alter the risk of lung cancer: a case-control analysis. *Lung Cancer* 2012;78:133-37.
27. Bodmer M, Becker C, Meier C, Jick SS, Meier CR. Use of metformin is not associated with a decreased risk of colorectal cancer: a case-control analysis. *Cancer Epidemiol Biomarkers Prev* 2012;21:280-86.
28. Bodmer M, Becker C, Meier C, Jick SS, Meier CR. Use of antidiabetic agents and the risk of pancreatic cancer: a case-control analysis. *Am J Gastroenterol* 2012;107:620-26.
29. Bodmer M, Becker C, Meier CJick SS, Meier CR Use of metformin and the risk of ovarian cancer: a case-control analysis. *Gynecol Oncol* 2011;123:200-04.
30. Bodmer M, Meier C, Krahenbuhl S, Jick SS, Meier CR. Long-term metformin use is associated with decreased risk of breast cancer. *Diabetes Care* 2010;33:1304-08.
31. Bosco JL, Antonsen S, Sorensen HT, Pedersen L, Lash TL. Metformin and incident breast cancer among diabetic women: a population-based case-control study in Denmark. *Cancer Epidemiol Biomarkers Prev* 2011;20:101-11.
32. Chaiteerakij R, Yang JD, Harmsen WS *et al*. Risk factors for intrahepatic cholangiocarcinoma: association between metformin use and reduced cancer risk. *Hepatology* 2013;57:648-55.
33. Chen HP, Shieh JJ, Chang CC *et al*. Metformin decreases hepatocellular carcinoma risk in a dose-dependent manner: population-based and in vitro studies. *Gut* 2013;62:606-15.
34. Dabrowski M. Glycated hemoglobin, diabetes treatment and cancer risk in type 2 diabetes. A case-control study. *Ann Agric Environ Med* 2013;20:116-21.
35. Donadon V, Balbi M, Mas MD, Casarin P, Zanette G. Metformin and reduced risk of hepatocellular carcinoma in diabetic patients with chronic liver disease. *Liver Int* 2010;30:750-58.

36. Donadon V, Balbi M, Valent F, Avogaro A. Glycated hemoglobin and antidiabetic strategies as risk factors for hepatocellular carcinoma. *World J Gastroenterol* 2010;16:3025-32.
37. Li D, Yeung SC, Hassan MM, Konopleva M, Abbruzzese JL. Antidiabetic therapies affect risk of pancreatic cancer. *Gastroenterology* 2009;13:482-88.
38. Hassan MM, Curley SA, Li D *et al*. Association of diabetes duration and diabetes treatment with the risk of hepatocellular carcinoma. *Cancer* 2010;116:1938-46.
39. Margel D, Urbach D, Lipscombe LL *et al*. Association between metformin use and risk of prostate cancer and its grade. *J Natl Cancer Inst* 2013;105:1123-31.
40. Mazzone PJ, Rai H, Beukemann M, Xu M, Jain A, Sasidhar M. The effect of metformin and thiazolidinedione use on lung cancer in diabetics. *BMC Cancer* 2012;12:410.
41. Monami M, Colombi C, Balzi D *et al*. Metformin and cancer occurrence in insulin-treated type 2 diabetic patients. *Diabetes Care* 2011;34:129-31.
42. Monami M, Lamanna C, Balzi D, Marchionni N, Mannucci E. Sulphonylureas and cancer: a case-control study. *Acta Diabetol* 2009;46:279-84.
43. Smiechowski BB, Azoulay L, Yin H, Pollak MN, Suissa S. The use of metformin and the incidence of lung cancer in patients with type 2 diabetes. *Diabetes Care* 2013;36:124-29.
44. Wang SY, Chuang CS, Muo CH *et al*. Metformin and the incidence of cancer in patients with diabetes: a nested case-control study. *Diabetes Care* 2013;36:e155-56.
45. Buchs AE, Silverman BG. 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;60:1379-85.
46. Chiu CC, Huang CC, Chen YC *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. *Intern Med* 2013;52:939-46.
47. Currie CJ, Poole CD, Evans M, Peters JR, Morgan CL. Mortality and other important diabetes-related outcomes with insulin vs other antihyperglycemic therapies in type 2 diabetes. *J Clin Endocrinol Metab* 2013;98:668-77.
48. Ferrara A, Lewis JD, Quesenberry CP Jr *et al*. Cohort study of pioglitazone and cancer incidence in patients with diabetes. *Diabetes Care* 2011;34:923-29.
49. Geraldine N, Marc A, Carla T *et al*. Relation between diabetes, metformin treatment and the occurrence of malignancies in a Belgian primary care setting. *Diabetes Res Clin Pract* 2012;97:331-36.
50. Hense HW, Kajuter H, Wellmann J, Batzler WU. Cancer incidence in type 2 diabetes patients - first results from a feasibility study of the D2C cohort. *Diabetol Metab Syndr* 2011;3:15.
51. Home PD, Kahn SE, Jones NP *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;53:1838-45.
52. Lai SW, Chen PC, Liao KF, Muo CH, Lin CC, Sung FC. Risk of hepatocellular carcinoma in diabetic patients and risk reduction associated with anti-diabetic therapy: a population-based cohort study. *Am J Gastroenterol* 2012;107:46-52.
53. Lai SW, Liao KF, Chen PC, Tsai PY, Hsieh DP, Chen CC. Antidiabetes drugs correlate with decreased risk of lung cancer: a population-based observation in Taiwan. *Clin Lung Cancer* 2012;13:143-48.
54. Lee MS, Hsu CC, Wahlqvist ML, Tsai HN, Chang YH, Huang YC. 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.
55. Lehman DM, Lorenzo C, Hernandez J, Wang CP. Statin use as a moderator of metformin effect on risk for prostate cancer among type 2 diabetic patients. *Diabetes Care* 2012;35:1002-07.
56. Morden NE, Liu SK, Smith J, Mackenzie TA, Skinner J, Korc M. Further exploration of the relationship between insulin glargine and incident cancer: a retrospective cohort study of older Medicare patients. *Diabetes Care* 2011;34:1965-71.
57. Neumann A, Weill A, Ricordeau P, Fagot JP, Alla F, Allemand H. Pioglitazone and risk of bladder cancer among diabetic patients in France: a population-based cohort study. *Diabetologia* 2012;55:1953-62.
58. Oliveria SA, Koro CE, Ulcickas Yood M, Sowell M. Cancer incidence among patients treated with antidiabetic pharmacotherapy. *Diabetes Metab Syndr* 2008;2:47-57.
59. Qiu H, Rhoads GG, Berlin JA, Marcella SW, Demissie K. Initial metformin or sulphonylurea exposure and cancer occurrence among patients with type 2 diabetes mellitus. *Diabetes Obes Metab*. 2013;15:349-57.
60. Redaniel MT, Jeffreys M, May MT, Ben-Shlomo Y, Martin RM. 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;23:1785-95.
61. Tsilidis KK, Capothanassi D, Allen N *et al*. Metformin does not affect cancer risk: A cohort study in the UK Clinical Practice Research Datalink analyzed like an intention to treat trial. *Diabetes Care* 2014;37:2522-32.
62. Van Staa TP, Patel D, Gallagher AM, De Bruin ML. Glucose-lowering agents and the patterns of risk for cancer: A study with the General Practice Research Database and secondary care data. *Diabetologia* 2012;55:654-65.
63. Yang X, So WY, Ma RC *et al*. Low HDL cholesterol, metformin use, and cancer risk in type 2 diabetes: the Hong Kong Diabetes Registry. *Diabetes Care* 2011;34:375-80.
64. Morgan CL, Poole CD, Evans M, Barnett AH, Jenkins-Jones S, Currie CJ. What next after metformin? A retrospective evaluation of the outcome of second-line, glucose-lowering therapies in people with type 2 diabetes. *J Clin Endocrinol Metab* 2012;97:4605-12.
65. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol* 1995;142:1255-64.
66. Hernan MA, Alonso A, Logan R *et al*. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 2008;19:766-79.
67. HIV-Causal Collaboration; Ray M, Logan R, Sterne JA *et al*. The effect of combined antiretroviral therapy on the

overall mortality of HIV-infected individuals. *AIDS*. 2010;24:123-37.

68. Walker AS, Ford D, Gilks CF *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. *Lancet* 2010;375:1278-86.
69. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JA. Methods for dealing with time-dependent confounding. *Stat Med* 2013;32:1584-618.
70. Ko EM, Sturmer T, Hong JL, Castillo WC, Bae-Jump V, Funk MJ. Metformin and the risk of endometrial cancer: a population-based cohort study. *Gynecol Oncol* 2015;13:341-47.

APPENDIX 4 BIAS ASSESSMENT CRITERIA IN FULL

Case control studies:

Outcome Assessment	**Unlikely:** Well defined/validated diagnosis e.g. Read codes, hospital records etc. Somehow incorporates / accounts for latency period before which cancer is likely present but not diagnosed (i.e. outcome is measured at a time that is in the appropriate risk period relative to exposure). **Low risk:** definition of cancer diagnosis with potential for non differential misclassification. **Medium risk:** Possible that outcome may occur at a time that is not in the appropriate risk period relative to exposure (e.g. No lag period applied to allow for latency of cancer) **High risk:** Unclear or un-validated method of diagnosis with potential for differential misclassification between exposed and unexposed. Outcomes will be occurring at a time that is not in the appropriate risk period relative to exposure.
Exposure Assessment	**Unlikely:** Well defined information e.g. prescription records, with level of exposure taken into account in a time dependent manner with at least 3 categories (low, medium, high / short term, medium term, long term/ current, past, never etc.) Or, if ITT analysis, minimum exposure period before being classified as "exposed" to ensure actually exposed. Choice of referent group made in consideration of confounding by indication, If cohort not incident users then past exposures should be measured and adjusted for. **Low risk:** As above but: less detailed measurement of cumulative exposure in a time varying analysis OR no minimum exposure period in ITT analysis OR some potential for choice of referent group to cause bias if appropriate adjustments not made elsewhere. OR some potential for misclassification of exposure due to lacking information on exposure prior to cohort entry. **Medium risk:** exposure assessed by patient recall OR strong potential for misclassification of exposure status due to lack of information about exposure prior to cohort entry (e.g. Assessing exposure in a small time interval (<6 months) around cohort entry in a prevalent user cohort). OR Strong potential for referent group to cause bias if appropriate adjustments not made. **High risk**: How exposure ascertained unknown/not clearly defined OR Future information used to inform exposure status at baseline.
Treatment of potential time dependent confounders (HBA1c, BMI, BP, Other medication)	**Unlikely:** Treated as time dependent confounders with appropriate methods (e.g. MSM's with IPTW or g computation) used to appropriately adjust. OR Time invariant exposure with confounder measured within a window prior to but close enough to exposure for its value to accurately represent level at time of exposure. Confounders measured as continuous or with sufficient detail (at least 3 categories). **Low: Fixed exposure** Adjusted for values measured before exposure but possibly a long time before actual exposure OR as unlikely, but with less detail in confounder measures OR Time varying exposure adjusted for baseline values of time dependent confounders. **Medium:** confounder measured at time point **after** exposure OR measured before exposure as presence of other things that may indicate level of disease severity or weight (i.e. Less accurate). OR as low but only some of the key TDCs adjusted for. **High:** None of the key TDCs considered in any way.

Baseline Adjustments	**Unlikely:** Adjusted for minimum of age, gender, some measure of disease severity (e.g. duration, previous treatment, HBA1c) BMI and smoking status (for all cancer and site specific known to be associated with smoking) with several categories for each (at least 3) or continuous if appropriate. Measured at correct time – i.e. baseline if it has potential to change through follow up. Measured from accurate information e.g. computer records. **Low:** Adjusted for all confounders but with less detail (e.g. Binary) OR adjustment for most of the minimum set with adequate detail and appropriate timing, but some variables from the minimum set omitted. **Medium:** Adjusted for most of the minimum but with less detail (e.g. Binary) or severity measure of any detail level but measured after exposure assessment/as average through follow up/at index date OR measured from inaccurate data e.g. Recall. **High:** Adjusted for age and gender only
Missing data	**Unlikely:** None or very low percentage of missing data, or appropriate missing data technique used for example Multiple imputation). Sensitivity analysis performed to assess potential impact of missingness. **Low:** Small amount of missing data (<15%) with no or inappropriate method applied to deal with it. **Medium:** Substantial missing data (15-25%) with no or inappropriate method used. **High:** Large amount of missing data (>25%) with no discussion /attempt to assess impact, or inappropriate method used.
Immortal Time	**Unlikely:** No follow up time included in which the event cannot occur by definition. **Low:** Immortal time included but not differential between exposure status **Medium:** differential immortal time between exposure groups but only likely to have small impact on results **High:** Immortal time in one exposure group but not another which is likely to have a significant impact on results.
Censoring	**Unlikely:** No censoring/loss to follow up present OR Some kind of sensitivity analysis or appropriate method of adjustment (e.g. IPW) used to assess impact of censoring. Censoring other than at loss to follow up (e.g. at change in treatment) is done with appropriate consideration of timing for exposure and outcome (e.g. not censoring at exact date of treatment change and hence excluding cancer event in the following week) if not accounted for in definition of exposure/outcome. **Low:** No specific method used to account for censoring but unlikely that censoring in study will have an impact on the results. **Medium:** No adjustments/additional analysis where it is possible that the censoring may cause bias OR potential for residual confounding between reason for censoring and outcome even after censoring adjustment applied. **High:** censoring present that will likely impact conclusions, with no discussion/analysis to assess impact.

Cohort studies:

Outcome Assessment	**Unlikely:** Well defined/validated diagnosis e.g. Read codes, hospital records etc. Somehow incorporates / accounts for latency period before which cancer is likely present but not diagnosed (i.e. outcome is measured at a time that is in the appropriate risk period relative to exposure). **Low risk:** definition of cancer diagnosis with potential for non differential misclassification. **Medium risk:** Possible that outcome may occur at a time that is not in the appropriate risk period relative to exposure (e.g. No lag period applied to allow for latency of cancer) **High risk:** Unclear or un-validated method of diagnosis with potential for differential misclassification between exposed and unexposed. Outcomes will be occurring at a time that is not in the appropriate risk period relative to exposure.

Exposure Assessment	**Unlikely:** Well defined information e.g. prescription records, with level of exposure taken into account in a time dependent manner with at least 3 categories (low, medium, high / short term, medium term, long term/ current, past, never etc.) Or, if ITT analysis, minimum exposure period before being classified as "exposed" to ensure actually exposed. Choice of referent group made in consideration of confounding by indication, If cohort not incident users then past exposures should be measured and adjusted for. **Low risk:** As above but: less detailed measurement of cumulative exposure in a time varying analysis OR no minimum exposure period in ITT analysis OR some potential for choice of referent group to cause bias if appropriate adjustments not made elsewhere. OR some potential for misclassification of exposure due to lacking information on exposure prior to cohort entry. **Medium risk:** exposure assessed by patient recall OR strong potential for misclassification of exposure status due to lack of information about exposure prior to cohort entry (e.g. Assessing exposure in a small time interval (<6 months) around cohort entry in a prevalent user cohort). OR Strong potential for referent group to cause bias if appropriate adjustments not made. **High risk**: How exposure ascertained unknown/not clearly defined OR Future information used to inform exposure status at baseline.
Treatment of potential time dependent confounders (HBA1c, BMI, BP, Other medication)	**Unlikely:** Treated as time dependent confounders with appropriate methods (e.g. MSM's with IPTW or g computation) used to appropriately adjust. OR Time invariant exposure with confounder measured within a window prior to but close enough to exposure for its value to accurately represent level at time of exposure. Confounders measured as continuous or with sufficient detail (at least 3 categories). **Low: Fixed exposure** Adjusted for values measured before exposure but possibly a long time before actual exposure OR as unlikely, but with less detail in confounder measures OR Time varying exposure adjusted for baseline values of time dependent confounders. **Medium:** confounder measured at time point **after** exposure OR measured before exposure as presence of other things that may indicate level of disease severity or weight (i.e. Less accurate). OR as low but only some of the key TDCs adjusted for. **High:** None of the key TDCs considered in any way.
Baseline Adjustments	**Unlikely:** Adjusted for minimum of age, gender, some measure of disease severity (e.g. duration, previous treatment, HBA1c) BMI and smoking status (for all cancer and site specific known to be associated with smoking) with several categories for each (at least 3) or continuous if appropriate. Measured at correct time – i.e. baseline if it has potential to change through follow up. Measured from accurate information e.g. computer records. **Low:** Adjusted for all confounders but with less detail (e.g. Binary) OR adjustment for most of the minimum set with adequate detail and appropriate timing, but some variables from the minimum set omitted. **Medium:** Adjusted for most of the minimum but with less detail (e.g. Binary) or severity measure of any detail level but measured after exposure assessment/as average through follow up/at index date OR measured from inaccurate data e.g. Recall. **High:** Adjusted for age and gender only
Missing data	**Unlikely:** None or very low percentage of missing data, or appropriate missing data technique used for example Multiple imputation). Sensitivity analysis performed to assess potential impact of missingness. **Low:** Small amount of missing data (<15%) with no or inappropriate method applied to deal with it. **Medium:** Substantial missing data (15-25%) with no or inappropriate method used. **High:** Large amount of missing data (>25%) with no discussion /attempt to assess impact, or inappropriate method used.
Immortal Time	**Unlikely:** No follow up time included in which the event cannot occur by definition. **Low:** Immortal time included but not differential between exposure status **Medium:** differential immortal time between exposure groups but only likely to have small impact on results **High:** Immortal time in one exposure group but not another which is likely to have a significant impact on results.

Censoring	**Unlikely:** No censoring/loss to follow up present OR Some kind of sensitivity analysis or appropriate method of adjustment (e.g. IPW) used to assess impact of censoring. Censoring other than at loss to follow up (e.g. at change in treatment) is done with appropriate consideration of timing for exposure and outcome (e.g. not censoring at exact date of treatment change and hence excluding cancer event in the following week) if not accounted for in definition of exposure/outcome. **Low:** No specific method used to account for censoring but unlikely that censoring in study will have an impact on the results. **Medium:** No adjustments/additional analysis where it is possible that the censoring may cause bias OR potential for residual confounding between reason for censoring and outcome even after censoring adjustment applied. **High:** censoring present that will likely impact conclusions, with no discussion/analysis to assess impact.

APPENDIX 5 STUDY LEVEL INFORMATION EXTRACTED

Case control studies										
Author	**Cancer site**	**Primary Relative Risk Estimate (95% CI)**	**Data Source**	**Exposure Definition**	**Comparator**	**Simplified Exposure Definition**	**Exposure Measurement time**	**Simplified Reference group**	**Incident users**	**Follow up time**
Azoulay (2011) [24]	Prostate	1.23 (0.99 , 1.52)	CPRD	Any prescription of metformin in the period between cohort entry and 1 year before index date	other OAD	Any Exposure	Single summary measure of exposure over entire follow up.	Any other OAD	Yes	Mean 4.7 years
Becker (2013) [25]	Ovarian/Endometrial	0.88 (0.58 , 1.32)	CPRD	long duration (>25 prescriptions) of metformin at least 2 years prior to index date	no metformin	Total Exposure (Number of prescriptions/time on metformin)	Single summary measure of exposure over entire follow up.	No metformin	No	Minimum 3 years
Bodmer (2010) [30]	Breast	1.09 (0.76 , 1.55)	CPRD	10 - 39 prescriptions of metformin prior to index date	no metformin	Total Exposure (Number of prescriptions/time on metformin)	Single summary measure of exposure over entire follow up.	No metformin	No	Minimum 3 years
Bodmer (2011) [29]	Ovarian/Endometrial	0.38 (0.14, 0.97)	CPRD	10-29 prescriptions of metformin prior to index date	no metformin	Total Exposure (Number of prescriptions/time on metformin)	Single summary measure of exposure over entire follow up.	No metformin	No	Not reported
Bodmer (2012) (Colorectal) [27]	Colorectal/ Bowel	1.05 (0.84 , 1.33)	CPRD	10-29 prescriptions of metformin prior to index date	no metformin	Total Exposure (Number of prescriptions/time on metformin)	Single summary measure of exposure over entire follow up.	No metformin	No	Minimum 3 years
Bodmer (2012) (Lung) [26]	Lung	1.21 (0.97 , 1.50)	CPRD	15 - 39 prescriptions of metformin prior to index date	no metformin	Total Exposure (Number of prescriptions/time on metformin)	Single summary measure of exposure over entire follow up.	No metformin	No	Minimum 3 years
Bodmer (2012) (Pancreatic) [28]	Pancreatic Cancer	0.83 (0.57 , 1.21)	CPRD	> 30 prescriptions of metformin prior to index date (only result reported in diabetic only analysis)	no metformin	Total Exposure (Number of prescriptions/time on metformin)	Single summary measure of exposure over entire follow up.	No metformin	No	Not reported
Bosco (2011) [31]	Breast	0.78 (0.59 , 1.01)	Danish Medical Registry	Minimum 1 year use of metformin	other OAD	Any exposure but minimum time/number of prescriptions needed)	Single summary measure of exposure over entire follow up.	Any other OAD	No	Not reported
Chen (2013) [33]	HCC/ICC	0.79 (0.75 , 0.83)	Taiwan National Health Insurance Claims database	Ever use of metformin in 5 years prior to index date	no metformin	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	No	5 years
Chaiteerakij (2013) [32]	HCC/ICC	0.20 (0.10, 0.40)	Mayo Clinic, Rochester MN	Ever use of metformin in the previous year	no metformin	Any Exposure	Current use	No metformin	No	1 year

Dabrowski (2013) [34]	All Cancer	0.26 (0.09 , 0.77)	Diabetic outpatient clinic, Poland	use of metformin prior to index date	no metformin	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	No	Not reported
Donadon (2010) [35]	HCC/ICC	0.15 (0.04 , 0.51)	Pordenone General Hospital, Italy	use of metformin at index date	sulfonylurea	Any Exposure	Current use	Sulfonylurea	No	Not reported
Donadon (2010) - 2 [36]	HCC/ICC	0.15 (0.04 , 0.50)	Pordenone General Hospital, Italy	use of metformin at index date	sulfonylurea	Any Exposure	Current use	Sulfonylurea	No	Not reported
Li (2009) [37]	Pancreatic Cancer	0.38 (0.22 , 0.69)	University of Texas MD Anderson Cancer Centre. (MDACC)	Self-reported ever use of metformin prior to index date	no metformin	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	No	Not reported
Evans (2005) [7]	All Cancer	0.77 (0.64 , 0.92)	Diabetes Audit Research Tayside (DARTS) and linked dispensed prescription database (MEMO)	Any metformin between 1993 and index date	no metformin	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	Not sure	Not reported
Hassan (2012) [38]	HCC/ICC	0.30 (0.20 , 0.60)	University of Texas MD Anderson Cancer Centre. (MDACC)	Self-reported ever use of metformin prior to index date	no metformin	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	No	Not reported
Margel (2013) [39]	Prostate	0.95 (0.85-1.07)	Ontario Diabetes Database, Ontario Cancer Registry and Ontario Drug Benefit Database	Ever use of metformin prior to index date	no metformin	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	Yes	Mean 2.9 years
Mazzone (2012) [40]	Lung	0.48 (0.28 , 0.81)	Diabetic Clinic, Cleveland USA	Metformin monotherapy only before index date	neither metformin nor thiazodaline used prior to index date	Monotherapy with metformin	Single summary measure of exposure over entire follow up.	No metformin	No	not reported
Monami (2009) [42]	All Cancer	0.60 (NO CI reported)	Diabetic outpatient clinic, University of Florence.	Any exposure to metformin in 10 years prior to index date	no metformin	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	No	Mean 6.5 years
Monami (2011) [41]	All Cancer	0.46 (0.25 , 0.85)	Diabetic outpatient clinic, University of Florence.	Ever exposure to metformin after start of insulin and before index date	no metformin	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	No	Mean 6.3 years
Smiechowski (2013) [43]	Lung	0.94 (0.76 , 1.17)	CPRD	Any prescription of metformin in the period between cohort entry (first ever prescription) and 1 year before index date	other OAD	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	Yes	Mean 5 years
Wang (2013) [44]	All Cancer Breast	1.06 (0.89 , 1.26) 1.05 (0.70 , 1.55)		Ever use of metformin prior to index date	no metformin	Any Exposure		No metformin	Not sure	Not reported

	Colorectal/ Bowel Lung Pancreas Prostate Stomach	0.94 (0.73 , 1.21) 1.11 (0.94 , 1.47) 1.14 (0.68 , 1.91) 0.94 (0.61 , 1.46) 1.62 (0.99 , 2.64)	Taiwan National Health Insurance Claims database				Single summary measure of exposure over entire follow up.			

Cohort Studies										
Author	**Cancer site**	**Primary Relative Risk Estimate (95% CI)**	**Data Source**	**Exposure Definition**	**Comparator**	**Simplified Exposure Definition**	**Exposure Measurement time**	**Simplified Reference group**	**Incident users**	**Follow up time**
Buchs (2011) [45]	All Cancer	1.00 (0.99 , 1.00)	Maccabi Healthcare Services (MHS) computerised databases	Cumulative use of metformin between 2003 & 2007	1 less prescription for metformin	Total Exposure (Number of prescriptions/time on metformin)	Single summary measure of exposure over entire follow up.	Less exposure	No	Mean 4.5 years
Chiu (2013) [46]	Colorectal/ Bowel Oesophagus Liver Pancreas Stomach	0.73 (0.43 , 1.24) 1.33 (0.49 , 3.59) 0.98 (0.74 , 1.29) 1.12 (0.63 , 2.00) 1.28 (0.72 , 2.08)	Taiwan National Health Insurance Claims database	Any exposure to metformin during the study period	no metformin	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	Not sure	Not reported
Currie (2009) [3]	All Cancer Breast Colorectal/Bowel Pancreas Prostate	1.36 (1.19 , 1.54) 0.98 (0.69 , 1.41) 1.80 (1.26 , 2.53) 4.95 (2.74 , 8.96) 1.07 (0.76 , 1.49)	The Health Information Network (THIN)	Newly initiated metformin monotherapy for at least 6 prescriptions	at least 6 prescriptions of sulfonylurea	Monotherapy with metformin for a minimum specified period	Time updated (current/ever/cumulative)	Sulfonylurea	Yes	Mean 2.3 years
Currie (2013) [47]	All Cancer	1.10 (1.00 , 1.20)	CPRD	Newly initiated metformin monotherapy for at least 180 days	sulfonylurea for at least 180 days	Monotherapy with metformin for a minimum specified period	Time updated (current/ever/cumulative)	Sulfonylurea	Yes	Mean 2.8 years
Ferrara (2011) [48]	Breast Colorectal/Bowel Kidney Lung	0.90 (0.80, 1.00) 1.00 (0.90 , 1.20) 1.30 (1.00 , 1.60) 1.00 (0.80 , 1.10)	Kaiser Perkanente Northern California Diabetes Registry (KPNC Diabetes registry)	2 or more prescriptions of metformin in a 6 month period needed to define ever use.	no metformin	Any exposure but minimum time/number of prescriptions needed)	Time updated (current/ever/cumulative)	No metformin	No	Mean 3.6 years

	Melanoma Pancreas Prostate	0.80 (0.60 , 1.10) 1.20 (1.00 , 1.50) 1.00 (0.90 , 1.10)								
Geraldine (2012) [49]	All Cancer	0.23 (0.09 , 0.60)	Intego database - Belgian primary care	Any exposure to metformin during the study period	diet only	Any Exposure	Single summary measure of exposure over entire follow up.	Diet	No	Mean 5 years
Hense (2011) [50]	All Cancer	0.95 (0.90 , 1.01)	German D2C cohort	Metformin monotherapy at cohort entry	no metformin	Monotherapy with metformin	Fixed from start of follow up, with exposure occurring in a baseline period or follow up starting from first exposure	No metformin	No	Mean 3.5 years
Home (2010) - Adopt [51]	All Cancer	0.92 (0.63 , 1.35)	ADOPT Clinical Trial	Randomisation to Metformin and Sulfonylurea	Randomisation to Rosiglitazone and Sulfonylurea	Randomisation to Metformin	Fixed from start of follow up, with exposure occurring in a baseline period or follow up starting from first exposure	Rosiglitazone	No	Mean 3.4 years
Home (2010) -Record [51]	All Cancer	1.22 (0.86 , 1.74)	RECORD Clinical Trial	Randomisation to Metformin and Sulfonylurea	Randomisation to Rosiglitazone and Sulfonylurea	Randomisation to Metformin	Fixed from start of follow up, with exposure occurring in a baseline period or follow up starting from first exposure	Rosiglitazone	#N/A	Mean 5.5 years
Hsieh (2012) [8]	All Cancer Breast Colorectal/Bowel Liver Lung Pancreas Prostate Stomach	0.56 (0.44 , 0.71) 0.57 (0.33 , 0.97) 0.54 (0.39 , 0.76) 0.66 (0.49 , 0.91) 0.64 (0.45 , 0.90) 0.63 (0.28 , 1.42) 0.97 (0.60 , 1.55) 0.65 (0.39 , 1.08)	Taiwan National Health Insurance Claims database	Metformin monotherapy for at least 1 year during study period	Sulfonylurea	Monotherapy for a minimum specified period	Single summary measure of exposure over entire follow up.	Sulfonylurea	No	Not reported
Lai (2012) (HCC) [52]	HCC/ICC	0.49 (0.37 , 0.66)	Taiwan National Health Insurance Claims database	Any exposure to metformin during the study period	other oad	Any Exposure	Single summary measure of exposure over entire follow up.	Any other OAD	Not sure	Mean 5.28 years
Lai (2012) (Lung) [53]	Lung	0.55 (0.37 , 0.82)	Taiwan National Health Insurance Claims database	Any exposure to metformin during the study period	other oad	Any Exposure	Single summary measure of exposure over entire follow up.	Any other OAD	Not sure	Mean 4.52 years
Lee (2011) [54]	Colorectal/Bowel Oesophagus Liver Pancreas Stomach	0.36 (0.13 , 0.98) 0.44 (0.07 , 2.61) 0.06 (0.02 , 0.16) 0.15 (0.03 , 0.79) 1.41 (0.42 , 4.73)	Taiwan National Health Insurance Claims database	At least 2 prescriptions of metformin during study period to define ever exposure, but also measured overall duration to metformin and adjusted for this in the final model	other oad	Total Exposure (Number of prescriptions/time on metformin)	Time updated (current/ever/cumulative)	Any other OAD	Yes	Mean 3.8 years

Lehman (2012) [55]	Prostate	1.68 (1.46 , 1.94)	Veteran Health Administration Health Care System	At least 180 days of metformin monotherapy	Sulfonylurea	Monotherapy for a minimum specified period	Fixed from start of follow up, with exposure occurring in a baseline period or follow up starting from first exposure	Sulfonylurea	No	Mean 5.2 years
Libby (2009) [5]	All Cancer Breast Colorectal/Bowel Lung	0.63 (0.53 , 0.75) 0.60 (0.32 , 1.10) 0.60 (0.38 , 0.94) 0.70 (0.43 , 1.15)	DARTS,Tayside Medicines Monitoring Unit, Scottish Morbidity record	More than 1 prescription of metformin in study period	no metformin	Any Exposure	Fixed from start of follow up, with exposure occurring in a baseline period or follow up starting from first exposure	No metformin	No	Not reported
Morden (2011) [56]	All Cancer Breast Colorectal/Bowel Pancreas Prostate	1.01 (0.94 , 1.08) 1.28 (1.05 , 1.57) 0.94 (0.72 , 1.22) 1.25 (0.89 , 1.75) 0.97 (0.76 , 1.24)	Medicare	Ever exposure to metformin in first 4 months after study enrolment	no metformin	Any Exposure	Fixed from start of follow up, with exposure occurring in a baseline period or follow up starting from first exposure	no met	No	Mean 1.93 years
Morgan (2012) [64]	All Cancer	0.93 (0.76 , 1.16)	GPRD	Metformin & Sulfonylurea combination therapy as 2nd line treatment after metformin monotherapy	Sulfonylurea only as 2nd line treatment after metformin monotherapy	Metformin + Sulfonylurea combination therapy	Fixed from start of follow up, with exposure occurring in a baseline period or follow up starting from first exposure	Sulfonylurea	No	Mean 2.3 years
Neumann (2011) [57]	Bladder Breast Colorectal/Bowel Kidney Lung	1.03 (0.93 , 1.14) 0.92 (0.88 , 0.97) 1.02 (0.98 , 1.07) 0.97 (0.89 , 1.05) 0.88 (0.84 , 0.92)	French national health insurance system - SBIIRAM (reimbursement database) and PMSI (hospital records) databases	At least 2 prescriptions of metformin over 6 consecutive months	other oad	Any exposure but minimum time/number of prescriptions needed)	Time updated (either no to yes or cumulative)	Any other OAD	No	Mean 3.1 years
Oliviera (2008) [58]	Bladder Colorectal/Bowel Liver Pancreas	0.99 (0.70 , 1.39) 0.67 (0.52 , 0.85) 0.73 (0.34 , 1.56) 1.26 (0.80 , 1.99)	US insurance database	Ever use of metformin during follow up	no met	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	No	Mean 3.9 years
Qiu (2013)	All Cancer	0.93 (0.86 , 1.02)	CPRD	at least 6 sequential prescriptions of metformin monotherapy	Sulfonylurea monotherapy	Monotherapy for a minimum specified period	Fixed from start of follow up, with exposure occurring in a baseline period or follow up starting from first exposure	Sulfonylurea	Not sure	Mean 3.8 years

Redaniel (2012) [60]	Breast	1.04 (0.79 , 1.37)	CPRD	Monotherapy with metformin for at least 6 months	Sulfonylurea monotherapy	Monotherapy for a minimum specified period	Time updated (either no to yes or cumulative)	Sulfonylurea	Yes	Mean 4.96 years
Ruiter (2012) [9]	All Cancer Breast Colorectal/Bowel Oesophagus HCC/ICC Lung Pancreas Prostate Stomach	0.90 (0.88 , 0.91) 0.95 (0.91 , 0.98) 0.91 (0.88 , 0.94) 0.90 (0.82 , 0.97) 0.67 (0.53 , 0.86) 0.87 (0.84 , 0.91) 0.73 (0.66 , 0.80) 0.92 (0.88 , 0.97) 0.83 (0.76 , 0.90)	PHARMO Record Linkage System (dispensing records from community pharmacies linked to hospital discharge records in Netherlands)	Monotherapy with metformin adjusting for cumulative use	Monotherapy with metformin adjusting for cumulative use	Total Exposure (Number of prescriptions/time on metformin)	Time updated (either no to yes or cumulative)	Sulfonylurea	Yes	Mean 3.5 years
Tsilidis (2014) [61]	All Cancer Bladder Breast Colorectal/Bowel Oesophagus Liver Leukaemia Lung Melanoma Ovarian/Endometrial Pancreas Prostate	0.96 (0.89 , 1.04) 0.88 (0.64 , 1.21) 1.03 (0.82 , 1.31) 0.92 (0.76 , 1.13) 1.05 (0.71 , 1.56) 0.85 (0.49 , 1.47) 0.86 (0.57 , 1.31) 0.85 (0.68 , 1.07) 1.26 (0.82 , 1.95) 1.38 (0.74 , 2.57) 0.70 (0.45 , 1.07) 1.02 (0.83 , 1.25)	CPRD	New users of metformin monotherapy	New users of sulfonylurea monotherapy	Monotherapy for a minimum specified period	Fixed from start of follow up, with exposure occurring in a baseline period or follow up starting from first exposure	Sulfonylurea	Yes	Mean 5.1 years
Van Staa (2012) [62]	All Cancer Breast Colorectal/Bowel Lung Pancreas Prostate	0.82 (0.75 , 0.90) 0.82 (0.61 , 1.10) 0.96 (0.71 , 1.29) 1.06 (0.80 , 1.41) 0.11 (0.07 , 0.16) 0.69 (0.52 , 0.91)	CPRD	At least 60 months exposure to metformin	0-6 months exposure to metformin	Total Exposure (Number of prescriptions/time on metformin)	Time updated (either no to yes or cumulative)	Less exposure	Yes	Mean 4.4 years
Yang (2011) [63]	All Cancer	0.45 (0.29 , 0.68)	Hong Kong Diabetes Registry	At least 1 prescription of metformin	no metformin	Any Exposure	Single summary measure of exposure over entire follow up.	No metformin	Not sure	Mean 5.5 years

APPENDIX 6 FULL BIAS ASSESSMENT

Study	Case (Outcome) Definition	Control Selection	Exposure Definition	Treatment of HBA1c, BMI, Other meds	Other Baseline Adjustments	Missing data
Azoulay (2011)	unlikely	unlikely	unlikely	Medium - BMI, HbA1c and other meds are nearest measure to index date, with ever exposure measured before this. Likely that this will include measures of HbA1c and other medications on causal pathway between metformin treatment and cancer rather than adjust for confounding.	Medium - Covariates measured at index date not at cohort entry, which may not correctly adjust for differences at cohort entry. Smoking only binary which potentially lacks detail. No information on duration of diabetes prior to starting therapy.	Unlikely/unknown - percentage of missing covariates generally low (reported for BMI and smoking) unknown percentage of missing hba1c though.
Becker (2013)	unlikely	unknown - not clear whether controls were cancer free at matched index date or cancer free for entire follow up (the latter having potential to induce bias).	low - not incident diabetic/user cohort so potential for miss classification of use	Medium - No adjustment for HBA1C (as didn't alter estimate when tested), adjustment for BMI and other OADs, but not clear when this BMI is measured relative to cohort entry or start of exposure. Likely to be on causal pathway rather than before treatment	Medium - Adjusted for diabetes duration but not clear how this was determined. Smoking measured with sufficient categorical detail. Matching did not take into account diabetes status, so in an analysis restricted to cases and controls with diabetes, it is not clear whether this matching was broken and/or whether matching factors were subsequently adjusted for.	unknown- Missing indicator method used for bmi and smoking but amount of missing data not reported (only reported for full cohort not diabetics only)
Bodmer (2011)	Low - although applied in a sensitivity analysis, when looking at diabetics only, potential latency of cancer not considered.	unknown - not clear whether controls were cancer free at matched index date or cancer free for entire follow up (the latter having potential to induce bias).	low - not incident diabetic/user cohort so potential for miss classification of use	Medium - adjustment for BMI but not clear when measured relative to cohort entry or start of exposure. Last recorded HbA1c before index date included, unlikely to correctly adjust for confounding by indication at time of exposure.	Medium - Adjusted for diabetes duration but not clear how this was determined. Smoking measured with sufficient categorical detail. Matching did not take into account diabetes status, so in an analysis restricted to cases and controls with diabetes, it is not clear whether this matching was broken and/or whether matching factors were subsequently adjusted for.	unknown - Missing indicator method used for bmi and smoking but amount of missing data not reported (only reported for full cohort not diabetics only)

Bodmer (2010)	Medium - no adjustment applied to allow for potential latency of cancer. Cancer must have occurred after use of OHA, however table 2 suggests patients on no treatment at all can be included. This means that cancers that should be attributed to no treatment (which contribute to the no metformin group) will be excluded disproportionately just because we know they later go on to use OHAs.	unlikely	medium - not matched on time in GPRD therefore potential that time for exposure ascertainment could be different in cases and controls - also, not incident users	Medium - adjustment for BMI but not clear when measured relative to cohort entry or start of exposure. Last recorded HbA1c before index date included, unlikely to correctly adjust for confounding by indication at time of exposure.	Low- Adjusted for age, gender, smoking and diabetes duration but not clear how the latter was determined - smoking measured with sufficient categorical detail.	unknown - Missing indicator method used for bmi and smoking but amount of missing data not reported (only reported for full cohort not diabetics only)
Bodmer (2012) (pancreatic)	Low - index date shifted back in primary analysis but not clear whether this shift was retained when analysis restricted to patients with diabetes only.	unknown - not clear whether controls were cancer free at matched index date or cancer free for entire follow up (the latter having potential to induce bias).	low - not incident diabetic/user cohort so potential for miss classification of use	Medium - No adjustment for HBA1C, adjustment for BMI and other OADs, but not clear when this BMI is measured relative to cohort entry or start of exposure. Likely to be on causal pathway rather than before treatment	Medium - Adjusted for diabetes duration but not clear how this was determined. Smoking measured with sufficient categorical detail. Matching did not take into account diabetes status, so in an analysis restricted to cases and controls with diabetes, it is not clear whether this matching was broken and/or whether matching factors were subsequently adjusted for.	unknown - Missing indicator method used for bmi and smoking but amount of missing data not reported (only reported for full cohort not diabetics only)
Bodmer (2012) (Colorectal cancer)	unlikely	unknown - not clear whether controls were cancer free at matched index date or cancer free for entire follow up (the latter having potential to induce bias).	low - not incident diabetic/user cohort so potential for miss classification of use	Medium - No adjustment for HBA1C (as didn't alter estimate when tested), adjustment for BMI and other OADs, but not clear when this BMI is measured relative to cohort entry or start of exposure. Likely to be on causal pathway rather than before treatment	Low - Adjusted for age, gender, and smoking and diabetes duration but not clear how the latter was determined or where measured with respect to exposure - smoking measured with sufficient categorical detail.	Low - <15% missing for smoking and bmi with missing indicator method used. Sensitivity analysis performed to look at the impact of missingness on relative risk estimates but not clear what.

Bodmer (2012) (Lung cancer)	unlikely	unknown - not clear whether controls were cancer free at matched index date or cancer free for entire follow up (the latter having potential to induce bias).	low - not incident diabetic/user cohort so potential for miss classification of use	Medium - No adjustment for HBA1C, adjustment for BMI and other OADs, but not clear when this BMI is measured relative to cohort entry or start of exposure. Likely to be on causal pathway rather than before treatment	Medium - Adjusted for diabetes duration but not clear how this was determined. Smoking measured with sufficient categorical detail. Matching did not take into account diabetes status, so in an analysis restricted to cases and controls with diabetes, it is not clear whether this matching was broken and/or whether matching factors were subsequently adjusted for.	unknown- Missing indicator method used for bmi and smoking but amount of missing data not reported (only reported for full cohort not diabetics only)
Bosco (2011)	Low - No lag applied to allow for potential latency of cancer diagnosis.	unlikely	Medium - "none exposed" comparator group very mixed and mix of incident and prevalent users- potential for confounding by disease severity. Recent/Past metformin user potentially useful but no measure of overall length/strength of exposure after the 1 year minimum. Not matched on time in database so potential for differing time windows between cases and control.	Medium - no adjustment made for use of other drugs within those who were exposed to metformin. BMI proxy of clinical obesity measured as occurring at all during follow up, Blood pressure and HbA1c not measured.	Medium - smoking not adjusted for. Other adjustments made - "diabetes complications" as ever/never during follow up. All measured between diagnosis and index date so not clear how measurement relates to timing of exposure	unlikely

Chaiteerakij (2013)	Low - No lag applied to allow for potential latency of cancer diagnosis.	low - although hospital based, care taken to ensure controls representative of population from which cases were taken and sensitivity analysis performed to look at effect of calendar time differences in recruitment. Possible that within diabetic patients, those who are able to visit the hospital for regular clinics and therefore enrolled as controls are generally healthier and therefore more likely to be on metformin, as opposed to more intensive treatments in those who have to come to the clinic for cancer treatment. Since no demographics are displayed for met vs no met or within diabetics only, this cannot be assessed.	High - exposure only measured as ever/never in the 1 year preceding diagnosis. Highly unlikely that exposure at this time would affect the development of cancer so quickly, and strong potential to consider those who have been previously exposed to metformin but are no longer on it (potentially because of disease progression) to be classified as not exposed.	High - no adjustments made at all.	High- matched on age and gender but nothing else	unlikely
Dabrowski (2013)	Low - No lag applied to allow for potential latency of cancer diagnosis.	Medium - controls cancer free for entire follow up but then are matched to the index date of the case.	High - appears to be current treatment rather than ever treatment. Strong potential for reverse causality. No measure of duration/dosage of metformin	High - only adjusted for use of other drugs, however how this is measured is not clear.	High- matched on age and gender but nothing else	low - 6/59 (10%) excluded because of missing HbA1c
Donadon (2010)	Low - No lag applied to allow for potential latency of cancer diagnosis.	Medium- Diabetic controls admitted for things other than Diabetes or liver related diseases, large proportion were admitted for heart failure and hypertension - commonly caused by being overweight? Therefore possible that they are more likely to be on metformin and not a sulfonylurea (which can cause further weight gain)	High- measured at admission/cancer diagnosis. Strong potential for reverse causality. Only measured as yes/no with no idea of duration or dosage of treatment.	Medium - Adjustment for A1c , BMI at time of diagnosis/enrolment only so likely on causal pathway	Medium - smoking status not adjusted for. Age > < 65 only,	Low - alcohol is only covariate with missing data and this was only approx. 4% in the cases only. Some data on diabetes treatment missing in both control groups but percentage very low.

Li (2009)	Low - No lag applied to allow for potential latency of cancer diagnosis.	Low - Controls selected from healthy relatives/spouses/friends accompanying cases to hospital appointments. Not clear whether this group is representative of the population of patients with diabetes as a whole who are at risk of pancreatic cancer.	Medium - Obtained from patient recall, so some potential for recall bias - those with more advanced stage diabetes who stopped taking metformin a long time ago may forget that they used it.	Medium - Only BMI adjusted for, and ever use of insulin but not other OADs. BMI taken as a mean of self-reported BMI at three time points, so may not reflect differences in BMI at time of exposure or time of diagnosis.	Medium - Diabetes duration/severity measured but not adjusted for in model. Smoking adjusted for but with minimum detail (ever/never) and all measured by recall.	low - numbers in tables suggest there is missing data for some covariates and also in terms of exposure status but proportions don't appear to be too large (not possible to work out exact missing amounts in patients with diabetes only)
Evans (2005)	Low - No lag applied to allow for potential latency of cancer diagnosis.	Medium - controls cancer free for entire follow up but then are matched to the index date of the case.	unlikely - with the exception of definition 1 (exposure in year before index date) which is unlikely to be a relevant risk period for a causal effect of metformin exposure on cancer risk	Medium - only BMI adjusted for, not clear at what time point it is measured. No info on other diabetes medications at all.	Low - all main baseline confounders included but not clear over what time interval they have been measured.	High - >25% missing for smoking, BMI and BP with missing category. Strong possibility for residual confounding.
Hassan (2012)	Low - No lag applied to allow for potential latency of cancer diagnosis.	Low - Controls selected from healthy relatives/spouses/friends accompanying cases to hospital appointments. Not clear whether this group is representative of the population of patients with diabetes as a whole who are at risk of pancreatic cancer.	Medium - assessed by patient recall and no time/level of exposure considered. Overall, duration of diabetes appears shorter in controls than cases therefore exposure time window may not be balanced between cases and controls	High - BMI, HBA1c and use of other drugs not considered at all.	Medium - smoking assessed from patient recall. Diabetes duration/severity not adjusted for	Unlikely - all information gathered by interview at one time point. Data requested unlikely to be unknown and therefore missing

Margel (2013)	Low - No lag applied to allow for potential latency of cancer diagnosis.	unlikely	Low - ever/never exposure potentially not representative of exposure that could actually affect risk of cancer, however cumulative duration definition better.	High - no adjustments made for HBA1c, BMI (unclear whether any of these co into the comorbidity index)	Unlikely - matched on diabetes duration which should account for baseline severity assuming measured well. Only missing adjustment is smoking status which may or may not be in the comorbidity index.	Unlikely - report relatively low percentage of missing covariate data and only for SES.
Mazzone (2012)	Low - No lag applied to allow for potential latency of cancer diagnosis.	Medium - controls cancer free for entire follow up but then are matched to the index date of the case.	Medium - duration only dichotomised as ever, or > 24 months. Matched on date of birth, but not time in database - if entire medical history of exposure not documented, this means there is potential for differing opportunities for exposure to be recorded between cases and controls. Also, not matched on length of diabetes diagnosis so quite possible that comparisons being made between patients at very different disease stages in terms of exposure.	Medium - BMI and A1c measured as mean through entire follow up. No adjustment for other medications such as sulfonylurea/insulin	Low - adjustments made but timing of measurements with respect to exposure is not clear. HbA1c is the only measure of "severity". Since not restricted to incident diabetes, adjustment for duration may also been useful.	unknown - Low levels of missing data for smoking, but not mentioned for other covariates

Monami (2009)	Low - No lag applied to allow for potential latency of cancer diagnosis.	Medium - controls cancer free for entire follow and recruited sequentially rather than at random.	Medium - mixture of medical records and self-reported exposure over the 10 years prior to cancer diagnosis/matched index date. Potential for recall bias.	Medium - matched on both BMI and HbA1c which were recorded at cohort entry. However, since patient recall of medication use before cohort entry used to inform exposure, this may not correct for confounding by indication, but may remove part of the total effect of metformin use on cancer risk. Only adjusted for ever use of other OHAs, timing of such a measurement makes interpretation unclear.	Unlikely - adjusted for duration of diabetes, smoking status age, gender and alcohol through matching	unlikely
Monami (2011)	Low - No lag applied to allow for potential latency of cancer diagnosis.	unlikely	Medium - well defined but low detail as only ever/never exposure to metformin. Only assessed exposure after insulin use so strong potential for misclassification of exposure to metformin. Some patient recall included.	Medium - bmi adjusted for by matching at insulin initiation. This does not necessarily precede metformin exposure since incident metformin use not established. No adjustment for HbA1c, and no adjustment for other oral OADS. Use of metformin or other OADs before insulin initiation was not considered.	Low- age and gender matched. Starting with incident insulin users will somewhat balance diabetes severity however CCS as a measure of severity of disease may not correctly adjust for differences between those who started metformin and not.	unlikely
Smiechowski (2013)	unlikely	unlikely	unlikely	Unlikely - cohort entry adjustments made for all TDCs and other medication entered as binary ever/never separately. Timing of measurement can be from 1 year before cohort entry up to 1 year before index date, therefore mixing baseline adjustment with adjustments that could potentially be on the causal pathway. A sensitivity analysis does address this and results are very similar.	Unlikely - all main baseline confounders included. Some potential to have adjusted for levels on causal pathway due to timing but sensitivity analysis performed to assess the impact of this.	unknown: approx. 18% missing HbA1c data for both cases and controls but not clear what ,if any, method used to deal with this

Wang (2013)	Low - No lag applied to allow for potential latency of cancer diagnosis.	Medium - controls cancer free until 2010 but then are matched to the index date of the case.	Low: Yearly median dose used in a secondary analysis but dichotomised only. Unknown how metformin use is actually established.	High: No adjustment for any potential TDC's	High - Only adjusted for age, gender and occupation - no adjustment for disease severity or smoking	Unlikely - not reported, but as an insurance database, age sex and occupation are unlikely to be missing.
Chen (2013)	Low - No lag applied to allow for potential latency of cancer diagnosis.	Low risk - those with previous liver surgery excluded, but was not applied to the cases. Not clear whether controls could have had a later HCC diagnosis	Low - observation period is only 5 years prior to index date, previous exposure not considered.	High- No adjustment for HbA1c or BMI. Use of other diabetic agents is included but measured during same window as total exposure so not clear whether adjustments are on causal pathway or not	Med- Age and gender matched, diabetes duration and severity (measured by number of visits) measured at index date. Severity in terms of number of visits could be affected by metformin use. Smoking not adjusted for.	unlikely - covariates used are yes no based on presence/absence of conditions, therefore not missing
Donadon (2010) - 2	Low - No lag applied to allow for potential latency of cancer diagnosis.	Medium- Diabetic controls admitted for things other than Diabetes or liver related diseases, large proportion were admitted for heart failure and hypertension - commonly caused by being overweight? Therefore possible that they are more likely to be on metformin and not a sulfonylurea (which can cause further weight gain)	High- measured at admission/cancer diagnosis. Strong potential for reverse causality. Only measured as yes/no with no idea of duration or dosage of treatment.	Medium - Adjustment for BMI at time of diagnosis/enrolment only. No adjustment for HbA1c and not clear how additional diabetes medications handled.	Low - smoking status not adjusted for, not entirely clear how diabetes duration was calculated.	unknown - no information provided

Study	Outcome Definition	Exposure Definition	Treatment of HBA1c, BMI, Other meds	Other Baseline Adjustments	Missing Data	Immortal Time	Censoring
Currie (2009)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - future information (remaining on medication for 180 days if intensified or 6 prescriptions if initial therapy) used to define exposure at baseline and at any subsequent treatment change. Will result in exclusion of some individuals who do not continue on medication, either because they do not survive or because they are intolerant, high risk of inducing a selection bias.	Med - None of the key TDCs adjusted for in final model due to lack of statistical significance but timing of measures questionable. Once in cohort 3 and 4, all have previous use of an OHA, but duration of previous exposure and type of drug unknown.	Medium - No adjustment for diabetes severity at baseline, no distinction between current/ex-smoker.	Unknown: no information on level of missing data for any covariates	Low: minimum exposure of 6 consecutive prescriptions to be included in the study however follow up starts at date of first prescription, therefore immortal time exists in the study, however this immortal time occurs for every medication and comparisons only made between medication groups. Unclear whether the immortal time is largely different between sulfonylurea and metformin (depends on prescription length) and so whether it will heavily bias the estimate.	High: censoring occurs at treatment change and if next (intensified) treatment doesn't last for 6 months they will be excluded from that cohort. Possible that reason for treatment not being continued related to outcome (e.g., they might already have cancer)
Currie (2013)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - future information (remaining on medication for 180 days) used to define exposure at baseline and at any subsequent treatment change. Will result in exclusion of some individuals who do not continue on medication, either because they do not survive or because they are intolerant, high risk of inducing a selection bias.	Low: Baseline adjustment made for HbA1c and BMI but does not consider previous use of OHAs which is necessary when time updating exposure between different groups.	Low: comprehensive list of potential baseline confounders but some with low detail - e.g. Smoking is only ever/never	unknown: multiple imputation used but no details given	Low: minimum exposure of 180 days but follow up starts from day 1. All cohort members must be on 1 drug so immortal follow up periods for everybody.	High: censoring occurs at treatment change and if next treatment doesn't last for 180 days months they will be excluded from that cohort. Possible that reason for treatment not being continued related to outcome (e.g., they might already have cancer)

Geraldine (2012)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - ever metformin use (so binary only) exposure allocated at baseline as fixed but is based on entire follow up period. No minimum period needed for metformin, single prescription will suffice, so further potential to classify people as exposed even if they do not continue with medication after first prescription.	Low- adjusted for at baseline (with the exception of combined/other diabetes medication) - weight and HBA1c entered as continuous	Low- measured at baseline and with good detail, but missing adjustment for diabetes duration which is important here due to combining prevalent and incident cases.	Unknown - levels of missing data not reported	High -exposure definition means all follow up time before exposure starts still apportioned as "exposed" time	Unknown - number lost to follow up/death not recorded
Home (2010)	Unlikely - no latency period but treatment randomised therefore decision to treat not affected by potential pre-existing cancer. Some potential for including cancers unrelated to any treatment, however this would not be differential between treatment groups.	unlikely	unlikely- randomised at baseline so able to estimate a total effect of initial treatment allocation (but this does not answer the question of the causal effect as didn't have 100% adherence)	Unlikely as treatment randomised	Unlikely	Unlikely	Low- RECORD <3% loss to follow up for unknown reasons. RECORD 526 subjects (out of over 4000) failed to adhere to monotherapy and were censored at this point which may have caused bias but the %age is quite low ADORT study - 21% non-adherence on metformin and 15% on rosiglitazone. Potential for a dilution of an effect but not likely to be huge?
Hsieh (2012)	Low - cancer in first year of follow up excluded to ensure cancer developed during study period, but first exposure could occur after this 1 year period.	High - Measured as monotherapy for at least 1 year on one of three treatments (for entire follow up). Excludes a large number of patients who will be on combination therapy at some point in their follow up - drop in n from 61,777 to 10,189 for the analysis suggests they have excluded anyone who has ever been on more than 1 drug or changed treatment. This choice of comparisons likely to be an a-typical group of patients - particularly those using insulin only. Cancer cases will have less follow up time to be exposed as if no cancer case the subject is followed up all the way until end of study period and logistic regression used.	High - not adjusted for at all.	High- adjusted for age and gender only.	Unlikely - not likely to have data on age and gender missing.	unlikely	Unknown - end of follow up classed as cancer diagnosis or 2008 (end of study). Not clear whether information on death / other censoring reasons available, but appear that the assumption made is if no cancer diagnosis then they are assumed alive and cancer free at end of study.

Lai (2012) (HCC)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - ever vs never exposure to metformin only, based on entire follow up period but assigned at baseline	Medium - No adjustment for use of other medications, no adjustment for HBA1c, though some baseline comorbidities adjusted for as yes/no presence (including obesity which could be a surrogate for BMI and partially adjust for severity of disease)	Medium - adjustment for age, gender but no smoking status and no other measure of disease severity. Due to how exposure defined, other adjustments (except age and gender) may not appropriately adjust for differences at time of treatment initiation. Insurance claims database so unhealthy behaviours likely to be unreported.	Unlikely - baseline covariate adjustments are yes/no for presence, where absence of information assumes no.	High -exposure definition means all follow up time before exposure starts still apportioned as "exposed" time	Unknown - number lost to follow up/death not recorded
Lai (2012) (LUNG)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - ever vs never exposure to metformin only, based on entire follow up period but assigned at baseline	Medium - No adjustment for use of other medications, no adjustment for HBA1c, though some baseline comorbidities adjusted for as yes/no presence (including obesity which could be a surrogate for BMI and partially adjust for severity of disease)	Medium - adjustment for age, gender, tobacco use and other comorbidity in form of propensity score, but no other measure of disease severity. Due to how exposure defined, other adjustments (except age and gender) may not appropriately adjust for differences at time of treatment initiation. Insurance claims database so unhealthy behaviours likely to be unreported.	Unlikely - baseline covariate adjustments are yes/no for presence, where absence of information assumes no.	High -exposure definition means all follow up time before exposure starts still apportioned as "exposed" time	Unknown - number lost to follow up/death not recorded

Lee (2011)	low - cancer only included if after 1 year of follow up from prescription, which will allow for latency of exposure effect/or of cancer diagnosis with regard to ever/never exposure to metformin, but may still be problematic when adjustments made for time updated duration of exposure.	High- follow up starts from first prescription of the drug for which they have been assigned based on knowledge of the entire follow up period. As such, the referent group is patients destined to never be on metformin, not those who may start on something else but end up on metformin. It is likely that these are people who have more severe disease and need to initiate straight onto stronger therapy, or those who have an absolute contraindication to metformin e.g. Renal disease, which may affect their underlying risk of certain cancers.	Med- adjusted for comorbidity score at baseline which may partially adjust for severity but not specifically BMI & HbA1c. Only adjustment for other medication is an ever/never exposure based on entire follow up - which by definition will always be "yes" in the referent group.	Medium - no adjustment for smoking status. Adjustment for "comorbidity" but this is general and not specific to disease severity. Age adjusted for with adequate detail.	Unlikely - baseline covariate adjustments are yes/no for presence, where absence of information assumes no.	Medium- Patients that start on an OAD other than metformin but move onto it later, have their survival time between first ever prescription and first metformin prescription censored. Therefore we systematically loose survival time which is cancer free (as they survive to go onto metformin) in that group, but cannot loose survival time in the metformin group.	low - only 2% loss to follow up (about the same in both metformin and referent group)
Libby (2009)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - When dosage included, it is maximum dose prescribed ever so potential for actual exposure levels for an individual to be miss classified. Choice of comparator group has potential to introduce bias due to selection process (non exposed selected from those who will never get metformin, not those who are not on metformin at matched index date) and allowing potential non exposed who would be excluded due to cancer before index date for one exposed subject to be matched to an exposed subject with a later metformin start date (therefore adding certain cancer events to the non exposed group).	Medium: using average of BMI and A1c throughout follow up likely to leave residual confounding by indication. Adjustment for use of other OADs at baseline will correctly adjust for baseline differences in previous medication use between those starting metformin and those not if the information on past use is accurate.	Unlikely - all key confounders measured. Aside from A1c and BMI (measured as means) they matched on year of diagnosis which as a diabetes registry may be more accurate than normal medical records	High: Unknown category included for smoking status, with 21 and 31% missing in exposed and unexposed respectively.	unlikely	Unknown: 34.8% of non-metformin users died from any cause as opposed to 14.9% of the metformin users, however it is not clear how many of these would have been censored due to death in the primary analysis.

Qui (2013)	unlikely- start of follow up is one year after first prescription, which ensures at least 1 year before outcome can occur (exposure defined as 6 consecutive prescriptions therefore it is possible that pre-existing cancer may have affected whether someone did or didn't have stable therapy in that year, but overall bias from this is unlikely).	Medium - defined by "stable treatment initiation" only, so not necessarily incident users (although incident users only was included in a sensitivity analysis), and we also know that they will have six consecutive prescriptions. This means selection bias possible since many people to do not continue on medication long enough may be excluded. Additionally, if the 6 prescriptions are over a time of >1 year then future information will still have been used to define exposure at baseline.	Low - all assessed and measured in year prior to baseline (start of treatment) but not included in final model for outcome of all cancer. HbA1c and Weight were measured as highest over a year period which may not be correct for levels at time of treatment initiation which is why low and not unlikely. Also, univariate vs single adjustment tests done to decide inclusion, which may not accurately show if something is a confounder within a full multivariable model, so questionable whether excluding from final model was reasonable.	Low - Age and gender only, measures of disease severity at baseline (duration, number of prescriptions in previous year) and smoking status were considered but not included. Low rather than unlikely due to system of excluding confounders	Med - substantial missing data for A1c (bit less for BMI). The sensitivity analysis performed suggested that missing A1c and diabetes duration did not change the unadjusted estimates by more than 10% for all malignancies, yet no further discussion or methods used to assess to what extent this could affect overall conclusions.	Low - minimum 6 consecutive prescriptions to be included as exposed, but this is true for all included. If average time for 6 sulfonylurea prescriptions is systematically different to 6 consecutive metformin prescriptions (and if one of these is >1 year) then risk time could be disproportionately affected.	Unlikely - sensitivity analysis occurred with extra censoring at 6 months past change from initial treatment (or end of study if that occurred before) and this had little effect on the results.
Redani el 2012	Unlikely - although not in primary analysis, exposure was re-examined including two different latency periods and results found to be similar, so care taken to insure bias not introduced by including outcomes in irrelevant risk periods.	Low - time varying exposure allows multiple cohort inclusion but within cohort no cumulative duration considered.	Medium -HbA1c adjusted for as an average, BMI adjusted for at baseline. No adjustment for use (even ever use) of other medication.	Med- Smoking not adjusted for. Age adjusted for with sufficient detail. HbA1c is only measure of disease severity and as mentioned in previous column this was done as an average which may not correctly adjust for confounding and may partially block causal pathway between exposure and outcome.	Unlikely. Multiple imputation used and well described.	Low - due to the condition for being "exposed" to a particular drug for > 6 months to be classed as exposed, there may be immortal time after entry to some categories that would not occur in others, however these categories are not included in the main analysis so only low risk. For example, if two patients start on sulfonylureas, and after 7 months one patient moves into the combination with metformin group, then we know that they must continue for 6 months more to have been classed as exposed to metformin as well as a sulfonylurea.	Unknown - loss to follow up due to death not reported. Sensitivity analysis of potential for ascertainment bias partly addresses issues with censoring directly at treatment change, as long latency periods are used (e.g.. 2/3 years which will account for more than just lag between actual development and diagnosis).

Ruiter (2012)	unlikely - different latency periods tested and no differences found	unlikely	High: HbA1c and BMI not considered as covariates at all. Use of other OADs not needed as an adjustment because they were an incident user group with ITT assumed and just cumulative use used as a time dependent covariate	Med: Number of hospitalisations used as a measure of comorbidity at baseline, but possible to have residual confounding from diabetes duration. Not able to adjust for smoking. I.e. Most of minimum set included (not smoking) but some with less than adequate detail.	Unknown - levels of missing data not reported	unlikely	Medium: censored at concomitant medication other than sulfs/metformin. Effect of exposure highly likely to be latent, so if there is a causal association, there is potential to miss cases by doing this.
Tsilidis (2014)	unlikely - used 12 month period after first exposure before follow up occurred, and changed to 6 months in sensitivity analysis	unlikely	Unlikely: No adjustment for baseline HbA1c in final model as included and found to not make much difference. Since incident group assuming ITT, this is reasonable to still estimate total effect of metformin vs sulf on cancer incidence. BMI adjusted for at baseline. Other medications not adjusted for but we assume ITT and then censor at treatment change from monotherapy. Baseline here is 1 year prior to index date, and HbA1c was a time weighted average so closest to time of initiation gets most weight therefore likely to be representative of A1c levels at time of treatment decision.	unlikely	Low - generally only small proportions of missing data (<15%) with the exception of alcohol consumption which is much higher (28%) with missing indicator method used (however alcohol not considered a key confounder so potential effect of this is probably small). No unadjusted effects reported, so unable to judge whether this method may leave residual confounding.	unlikely	Low - weighting applied to adjust for non-adherence but not for other reasons for censoring, but this only provides a valid adherence adjusted estimate if the model predicting adherence is correctly specified and there are no unmeasured causes of censoring and cancer. Since censoring applied on exact date of change of treatment, any change in treatment due to presence of pre-diagnosed cancer would not be accounted for. No information reported on numbers censored due to loss to follow up/death.

Yang (2011)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - ever/never metformin use based on entire follow up period but exposure allocated at baseline.	Low - all potential TDCs measured at cohort entry but since exposure is measured throughout follow up, and cohort entry is enrolment, the measures at this time point may not correctly adjust for differences at time of treatment initiation.	Unlikely - comprehensive list of baseline confounders that appear to be measured accurately at cohort entry (though slight uncertainty as to whether they are actually incident cases).	unknown - complete case analysis used but not clear how much missing data there was for covariates (only HDL cholesterol reported)	High -exposure definition means all follow up time before exposure starts still apportioned as "exposed" time	unknown - loss to follow up due to death not reported
Buchs (2011)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - Baseline period for exposure assessment from 2000 - 2002, with follow up exposure from 2003 onwards. However no requirement for included subjects to be newly diagnosed (just to have not had any insulin). Potential for bias induced by comparison of subjects at inherently different disease stages since period for exposure ascertainment somewhat arbitrary with respect to start of disease. Future information on dosage used to inform total dosage at baseline.	Medium - Only other medication adjusted for.	High- no adjustment for smoking or any measure of disease severity. Age dichotomised to < or > 65 only.	Unlikely - absence of measure would indicate a "no" rather than missing.	High - use of total number of prescriptions as a continuous predictor of exposure over all of follow up means those with more prescriptions could just be surviving to have them.	Medium - approx. 14% loss to follow up from health problems or leaving database. Therefore likely to have shorter exposure period and will not have a cancer diagnosis.
Chiu (2013)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - ever/never metformin use based on entire follow up period but exposure allocated at baseline.	High - not adjustment for HbA1c or BMI at all, not clear whether each medication comparison was modelled separately or whether all in one model , so possibly no adjustment for other medication usage either	Medium - no adjustment for smoking, there was adjustment for “selected comorbidities" but not clear when these were measured and there may still be residual confounding by specific diabetes severity.	Unlikely: absence of measure would indicate a "no" rather than missing.	High -exposure definition means all follow up time before exposure starts still apportioned as "exposed" time	Unknown - no information on loss to follow up because of death.

Ferrara (2011)	Low - initial 6 month or 12 month exclusion period does not exactly control for latency of cancer since exposure is time updated, and later changes in exposure could still be affected by pre-diagnosed cancer	Low - not incident users so may be some misclassification in terms of past exposure	Medium - All adjusted for baseline value with exception of BMI. As exposure is time dependent this may not correctly adjust for those who are treated later in time.	Low - most of the key variables adjusted for with the exception of diabetes duration which may be important since not an incident cohort.	Medium - approx. 20% missing HbA1c and approx. 18% ethnicity, with missing indicator method used	unlikely	Unknown - no information on loss to follow up because of death.
Olivieria (2008)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - only ever/never or current use vs current use of other drugs. Current/past definition is unclear and does not take into account latency effect - i.e. Even if someone stops taking metformin the effect on cancer will occur well beyond "past use" and as such not be included? Ever/Never exposure measured over entire follow up period as ever use so future information used to inform baseline exposure.	High - no adjustment at all	High - adjusted for age and sex only, nothing to assess disease severity	Unlikely - not likely to have data on age and gender missing.	High -exposure definition of ever/never means all follow up time before exposure starts still apportioned as "exposed" time	Unknown - number lost to follow up/death not recorded
Hense (2011)	Low - No adjustment to allow for potential latency of cancer diagnosis.	Medium - "medication at cohort entry" is current medication, so strong potential to miss people that have had prior exposure to medication - such people may be at differing stages of disease.	Low - adjusted for BMI (study entry) and other meds (study entry) but not HbA1c.	low - adjusted for all but smoking status	unlikely	unlikely	Medium - if no cancer record then assumed to have survived. This could potentially add extra survival time, particularly to those who are more severe (and have actually died of other causes) who would more likely be on multiple medications and therefore not on metformin monotherapy.

Lehman (2012)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - follow up starts from first metformin or sulfonylurea prescription, but definition requires 180 days of that prescription and no use of other drugs (TZD or insulin) and no more than 180 days of the comparator drug, for entire prescription period. High risk that this will induce selection bias by excluding people who do not fit this specific exposure pattern. Dose dependent analysis uses proportion of days at higher dose, which is relative not absolute so may not compare actual exposure levels correctly.	Medium- HbA1c measured as an average through follow up so not clear that this would adjust for differences at therapy initiation and could be adjusting for things on the causal pathway. No adjustment for BMI.	Low: diabetes duration only dichotomised, Smoking status appears to be defined as "on smoking cessation therapy" which may be less accurate.	Unknown: complete case analysis used, not clear what proportions of subjects were excluded due to missing data.	Low: minimum exposure of 180 days but fop starts from day 1. All cohort members must be on 1 drug so immortal follow up periods for everybody.	unknown - loss to follow up due to death not reported
Liao (2012)	Low - No adjustment to allow for potential latency of cancer diagnosis.	High - yes/no exposure to metformin during the follow up period so future information is being used to define exposure from baseline.	high: univariate analysis	high: univariate analysis	Unknown: no information on level of missing data for any covariates	Medium: immortal time between entry into cohort and time of medication included - everyone must be on an oad, so everyone will have time added but if time to medication differs between medication groups this will cause bias.	unknown - loss to follow up due to death/transfer from database not reported
Morden (2011)	Low - incident cases counted as those diagnosed after the 4 month exposure assessment window, but this does not ensure a minimum exposure of 4 months. Still potential for reverse causality.	Medium - all must have had an insulin prescription in the first 4 months of enrolment in the part D program. Meaning all have relatively advanced t2DM at entry. Metformin use only established within this window, and prior use not considered which may be important considering any effect would be expected to occur years after exposure.	Medium - HbA1c and BMI not considered at all but diabetes complications are used as an adjustment for severity which could be considered a proxy for HbA1c.	unlikely - key fixed confounders included and some proxy measures of severity	unknown - not clear how much missing data there was (though most covariates absence = no)	unlikely	unknown - loss to follow up not reported

Neumann (2011)	Low - No adjustment to allow for potential latency of cancer diagnosis. Since exposure is time updated, the 6 month "baseline" exposure does not guarantee that the exposure status could not be affected by undiagnosed cancer.	Low- no exposure information prior to 2006 and they are not incident diabetic cases so potential for misclassification.	High - HbA1c and BMI not considered at all	Low - Age gender and proxy for diabetes duration included, data on smoking not available.	Unlikely - no missing data reported for variables used in analysis	unlikely	unknown - loss to follow up not reported
Van Staa (2012)	Low - multiple analyses to assess patterns of risk, some of which may be affected by the fact that no latency period used, but others will not be affected by this. Some potential for main estimate used to be affected.	unlikely	Low - measured at baseline, all key included except HbA1c.	Unlikely - all key baseline confounders included. Nothing specific for diabetes severity but all patients are new users of OADs. Other comorbidities may partially account for diabetes severity.	Low - MIM used for BMI as categorical but presented as continuous mean (SD) so extent missing unknown. Probably quite low based on knowledge of BMI recording for diabetics in CPRD. Unknown smoking status low proportion missing.	unlikely	low/med: for between treatment comparisons censored directly at change time so cancer early in new exposure unlikely to be attributable to new exposure but results relating to long term use comparisons may not have this problem.
Morgan (2012)	Low - No adjustment to allow for potential latency of cancer diagnosis.	unlikely	Unlikely - all TDCs measured at cohort entry, so will adjust for differences between those who continue with metformin and those who switch.	unlikely - all key adjustments made at baseline	Unknown - only 0.3% missing data on smoking status. But missingness not reported for HbA1c and BMI - though have used CPRD so bound to be some!	Low: minimum exposure of 90 days but fop starts from day 1. All cohort members must be on 1 drug so immortal follow up periods for everybody.	Medium - Censored at change in therapy - potential to miss cancer cases occurring after this point. Loss to follow up for this reason and death etc. not reported so unclear how this may affect it but certainly there is a risk

APPENDIX 7 CODELISTS FOR DIABETES DIAGNOSIS

Initial diagnosis codelist

medcode	readcode	readterm	code_cat
711	C10..00	Diabetes mellitus	Vague codes
758	C10F.00	Type 2 diabetes mellitus	Definite T2 codes
506	C100112	Non-insulin dependent diabetes mellitus	Possible T2 codes
4513	C109.00	Non-insulin dependent diabetes mellitus	Probable T2 codes
1549	C10E.00	Type 1 diabetes mellitus	Definite T1 codes
1410	C112.00	Hypoglycaemia unspecified	Possible T2 codes
10983	C11y300	Impaired fasting glycaemia	Possible T2 codes
17859	C109.12	Type 2 diabetes mellitus	Probable T2 codes
10921	C11y200	Impaired glucose tolerance	Possible T2 codes
1038	C100011	Insulin dependent diabetes mellitus	Possible T1 codes
1647	C108.00	Insulin dependent diabetes mellitus	Probable T1 codes
14889	C100111	Maturity onset diabetes	Possible T2 codes
8446	L180811	Gestational diabetes mellitus	Definite Gestational diabetes
14803	C100100	Diabetes mellitus, adult onset, no mention of complication	Possible T2 codes
1682	C101.00	Diabetes mellitus with ketoacidosis	Vague codes
1407	C10FJ00	Insulin treated Type 2 diabetes mellitus	Definite T2 codes
16230	C106.00	Diabetes mellitus with neurological manifestation	Vague codes
7795	C106.12	Diabetes mellitus with neuropathy	Vague codes
18505	C108.11	IDDM-Insulin dependent diabetes mellitus	Probable T1 codes
18390	C10FM00	Type 2 diabetes mellitus with persistent microalbuminuria	Definite T2 codes
17858	C108.12	Type 1 diabetes mellitus	Probable T1 codes
5884	C109.11	NIDDM - Non-insulin dependent diabetes mellitus	Probable T2 codes
18219	C109.13	Type II diabetes mellitus	Probable T2 codes
10692	C10EM00	Type 1 diabetes mellitus with ketoacidosis	Definite T1 codes
18278	C109J00	Insulin treated Type 2 diabetes mellitus	Probable T2 codes
26054	C10FL00	Type 2 diabetes mellitus with persistent proteinuria	Definite T2 codes
2472	C110.00	Hypoglycaemic coma	Possible T1 codes
2475	C104.11	Diabetic nephropathy	Vague codes
22884	C10F.11	Type II diabetes mellitus	Definite T2 codes
2664	L180900	Gestational diabetes mellitus	Definite Gestational diabetes
38986	C100.00	Diabetes mellitus with no mention of complication	Vague codes
12640	C10FC00	Type 2 diabetes mellitus with nephropathy	Definite T2 codes
4563	C112000	Reactive hypoglycaemia NOS	Possible T2 codes
8403	C109700	Non-insulin dependent diabetes mellitus - poor control	Probable T2 codes
35399	C107.00	Diabetes mellitus with peripheral circulatory disorder	Vague codes
18496	C10F600	Type 2 diabetes mellitus with retinopathy	Definite T2 codes
33254	C105.00	Diabetes mellitus with ophthalmic manifestation	Vague codes
16502	C104.00	Diabetes mellitus with renal manifestation	Vague codes
24490	C100000	Diabetes mellitus, juvenile type, no mention of complication	Possible T1 codes
20368	C112z00	Hypoglycaemia unspecified NOS	Possible T2 codes
25627	C10F700	Type 2 diabetes mellitus - poor control	Definite T2 codes
11551	C10B.00	Diabetes mellitus induced by steroids	Secondary / Other types
11359	L180.00	Diabetes mellitus during pregnancy/childbirth/puerperium	Probable Gestational diabetes
24423	C108.13	Type I diabetes mellitus	Probable T1 codes
47954	C10F900	Type 2 diabetes mellitus without complication	Definite T2 codes
32627	C10FN00	Type 2 diabetes mellitus with ketoacidosis	Definite T2 codes
51261	C10E.12	Insulin dependent diabetes mellitus	Definite T1 codes
50972	C100z00	Diabetes mellitus NOS with no mention of complication	Vague codes
31310	C108900	Insulin dependent diabetes maturity onset	Probable T1 codes
30323	C10EK00	Type 1 diabetes mellitus with persistent proteinuria	Definite T1 codes
10418	C10ED00	Type 1 diabetes mellitus with nephropathy	Definite T1 codes
6791	C108800	Insulin dependent diabetes mellitus - poor control	Probable T1 codes
18387	C10E700	Type 1 diabetes mellitus with retinopathy	Definite T1 codes
30294	C10EL00	Type 1 diabetes mellitus with persistent microalbuminuria	Definite T1 codes
34450	C10FK00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus	Definite T2 codes
10278	L180800	Diabetes mellitus arising in pregnancy	Definite Gestational diabetes
32885	C11y100	Drug-induced hypoglycaemia without coma	Secondary / Other types
15690	C103.00	Diabetes mellitus with ketoacidotic coma	Vague codes

42505	C101z00	Diabetes mellitus NOS with ketoacidosis	Vague codes
35107	C104z00	Diabetes mellitus with nephropathy NOS	Vague codes
32193	C11y000	Steroid induced diabetes	Secondary / Other types
36695	C10D.00	Diabetes mellitus autosomal dominant type 2	Genetic
32403	C107.11	Diabetes mellitus with gangrene	Vague codes
35288	C10E800	Type 1 diabetes mellitus - poor control	Definite T1 codes
29979	C109900	Non-insulin-dependent diabetes mellitus without complication	Probable T2 codes
22573	C106z00	Diabetes mellitus NOS with neurological manifestation	Vague codes
6509	C108700	Insulin dependent diabetes mellitus with retinopathy	Probable T1 codes
34912	C109400	Non-insulin dependent diabetes mellitus with ulcer	Probable T2 codes
50960	L180500	Pre-existing diabetes mellitus, insulin-dependent	Probable T1 codes
21482	C102.00	Diabetes mellitus with hyperosmolar coma	Vague codes
44443	C108500	Insulin dependent diabetes mellitus with ulcer	Probable T1 codes
12455	C10E.11	Type I diabetes mellitus	Definite T1 codes
40837	C10EN00	Type 1 diabetes mellitus with ketoacidotic coma	Definite T1 codes
39317	C106100	Diabetes mellitus, adult onset, + neurological manifestation	Possible T2 codes
51697	C10G.00	Secondary pancreatic diabetes mellitus	Secondary / Other types
18425	C10FB00	Type 2 diabetes mellitus with polyneuropathy	Definite T2 codes
34283	C105z00	Diabetes mellitus NOS with ophthalmic manifestation	Vague codes
53392	C10F911	Type II diabetes mellitus without complication	Definite T2 codes
25591	C10FQ00	Type 2 diabetes mellitus with exudative maculopathy	Definite T2 codes
18777	C10F000	Type 2 diabetes mellitus with renal complications	Definite T2 codes
35385	C10FH00	Type 2 diabetes mellitus with neuropathic arthropathy	Definite T2 codes
46624	C10C.11	Maturity onset diabetes in youth	Genetic
34268	C10F200	Type 2 diabetes mellitus with neurological complications	Definite T2 codes
39481	C10F811	Metabolic syndrome X	Not diabetes
24405	C112100	Spontaneous hypoglycaemia NOS	Possible T2 codes
55239	C10EQ00	Type 1 diabetes mellitus with gastroparesis	Definite T1 codes
44982	C10FE00	Type 2 diabetes mellitus with diabetic cataract	Definite T2 codes
64357	C10zz00	Diabetes mellitus NOS with unspecified complication	Vague codes
17262	C109600	Non-insulin-dependent diabetes mellitus with retinopathy	Probable T2 codes
26108	C10B000	Steroid induced diabetes mellitus without complication	Secondary / Other types
49074	C10F400	Type 2 diabetes mellitus with ulcer	Definite T2 codes
32556	C107.12	Diabetes with gangrene	Vague codes
63690	C10FR00	Type 2 diabetes mellitus with gastroparesis	Definite T2 codes
41389	C105100	Diabetes mellitus, adult onset, + ophthalmic manifestation	Possible T2 codes
64668	C10FJ11	Insulin treated Type II diabetes mellitus	Definite T2 codes
33969	C10A100	Malnutrition-related diabetes mellitus with ketoacidosis	Secondary / Other types
39070	C10EE00	Type 1 diabetes mellitus with hypoglycaemic coma	Definite T1 codes
53200	C101000	Diabetes mellitus, juvenile type, with ketoacidosis	Possible T1 codes
37957	C10K.00	Type A insulin resistance	Not diabetes
67853	C106000	Diabetes mellitus, juvenile, + neurological manifestation	Possible T1 codes
65025	C107z00	Diabetes mellitus NOS with peripheral circulatory disorder	Vague codes
47315	C10F711	Type II diabetes mellitus - poor control	Definite T2 codes
46917	C10FD00	Type 2 diabetes mellitus with hypoglycaemic coma	Definite T2 codes
45491	C10z.00	Diabetes mellitus with unspecified complication	Vague codes
16491	C106.13	Diabetes mellitus with polyneuropathy	Vague codes
26855	C108400	Unstable insulin dependent diabetes mellitus	Probable T1 codes
22487	C10N.00	Secondary diabetes mellitus	Secondary / Other types
33343	C10y.00	Diabetes mellitus with other specified manifestation	Vague codes
51371	C110z00	Hypoglycaemic coma NOS	Possible T2 codes
62674	C10FA00	Type 2 diabetes mellitus with mononeuropathy	Definite T2 codes
60796	C10FL11	Type II diabetes mellitus with persistent proteinuria	Definite T2 codes
22871	C10EP00	Type 1 diabetes mellitus with exudative maculopathy	Definite T1 codes
47321	C10F100	Type 2 diabetes mellitus with ophthalmic complications	Definite T2 codes
59365	C109C00	Non-insulin dependent diabetes mellitus with nephropathy	Probable T2 codes
93380	C10N100	Cystic fibrosis related diabetes mellitus	Secondary / Other types
95636	C10ER00	Latent autoimmune diabetes mellitus in adult	Definite T1 codes
63357	C107100	Diabetes mellitus, adult, + peripheral circulatory disorder	Possible T2 codes
63762	C10z100	Diabetes mellitus, adult onset, + unspecified complication	Probable T2 codes
69676	C10EA00	Type 1 diabetes mellitus without complication	Definite T1 codes
44440	C108E00	Insulin dependent diabetes mellitus with hypoglycaemic coma	Probable T1 codes
54008	C10EJ00	Type 1 diabetes mellitus with neuropathic arthropathy	Definite T1 codes
52236	C10A.00	Malnutrition-related diabetes mellitus	Secondary / Other types
34639	L180100	Diabetes mellitus during pregnancy - baby delivered	Probable Gestational diabetes
50609	L180600	Pre-existing diabetes mellitus, non-insulin-dependent	Probable T2 codes
12736	C10F500	Type 2 diabetes mellitus with gangrene	Definite T2 codes

18683	C10E500	Type 1 diabetes mellitus with ulcer	Definite T1 codes
35105	C104100	Diabetes mellitus, adult onset, with renal manifestation	Possible T2 codes
40682	C10E900	Type 1 diabetes mellitus maturity onset	Definite T1 codes
33807	C107200	Diabetes mellitus, adult with gangrene	Possible T2 codes
37806	C10FF00	Type 2 diabetes mellitus with peripheral angiopathy	Definite T2 codes
43921	C10E400	Unstable type 1 diabetes mellitus	Definite T1 codes
64384	L180z00	Diabetes mellitus in pregnancy/childbirth/puerperium NOS	Probable Gestational diabetes
18264	C109J12	Insulin treated Type II diabetes mellitus	Probable T2 codes
24458	C109711	Type II diabetes mellitus - poor control	Probable T2 codes
47582	C10E000	Type 1 diabetes mellitus with renal complications	Definite T1 codes
49559	L180300	Diabetes mellitus during pregnancy - baby not yet delivered	Probable Gestational diabetes
54856	C101100	Diabetes mellitus, adult onset, with ketoacidosis	Vague codes
45913	C109712	Type 2 diabetes mellitus - poor control	Probable T2 codes
36633	C109K00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus	Probable T2 codes
59253	C10FG00	Type 2 diabetes mellitus with arthropathy	Definite T2 codes
46963	C108000	Insulin-dependent diabetes mellitus with renal complications	Probable T1 codes
47650	C10E300	Type 1 diabetes mellitus with multiple complications	Definite T1 codes
39809	C108J00	Insulin dependent diab mell with neuropathic arthropathy	Probable T1 codes
49655	C10F611	Type II diabetes mellitus with retinopathy	Definite T2 codes
61577	C11y.00	Other specified disorders of pancreatic internal secretion	Secondary / Other types
65267	C10F300	Type 2 diabetes mellitus with multiple complications	Definite T2 codes
43139	C102100	Diabetes mellitus, adult onset, with hyperosmolar coma	Possible T2 codes
54773	C10F800	Reaven's syndrome	Not diabetes
43857	C10M.00	Lipoatrophic diabetes mellitus	Secondary / Other types
51756	C10FP00	Type 2 diabetes mellitus with ketoacidotic coma	Definite T2 codes
56803	C107400	NIDDM with peripheral circulatory disorder	Probable T2 codes
61122	C10H.00	Diabetes mellitus induced by non-steroid drugs	Secondary / Other types
43453	C10C.00	Diabetes mellitus autosomal dominant	Genetic
58604	C109611	Type II diabetes mellitus with retinopathy	Probable T2 codes
59991	C10D.11	Maturity onset diabetes in youth type 2	Genetic
62209	C10EM11	Type I diabetes mellitus with ketoacidosis	Definite T1 codes
42831	C10E200	Type 1 diabetes mellitus with neurological complications	Definite T1 codes
38617	C101y00	Other specified diabetes mellitus with ketoacidosis	Vague codes
46301	C10EC00	Type 1 diabetes mellitus with polyneuropathy	Definite T1 codes
55075	C109411	Type II diabetes mellitus with ulcer	Probable T2 codes
37648	C109J11	Insulin treated non-insulin dependent diabetes mellitus	Probable T2 codes
49554	C10EF00	Type 1 diabetes mellitus with diabetic cataract	Definite T1 codes
56448	C108A00	Insulin-dependent diabetes without complication	Probable T1 codes
39406	C109800	Reaven's syndrome	Not diabetes
46850	C108811	Type I diabetes mellitus - poor control	Probable T1 codes
57621	C108D00	Insulin dependent diabetes mellitus with nephropathy	Probable T1 codes
61523	C106y00	Other specified diabetes mellitus with neurological comps	Vague codes
50429	C109100	Non-insulin-dependent diabetes mellitus with ophthalm comps	Probable T2 codes
52303	C109000	Non-insulin-dependent diabetes mellitus with renal comps	Probable T2 codes
59903	C106.11	Diabetic amyotrophy	Vague codes
61520	C110000	Iatrogenic hyperinsulinism	Possible T2 codes
68390	C108512	Type 1 diabetes mellitus with ulcer	Probable T1 codes
42729	C108E11	Type I diabetes mellitus with hypoglycaemic coma	Probable T1 codes
42762	C109612	Type 2 diabetes mellitus with retinopathy	Probable T2 codes
55842	C109200	Non-insulin-dependent diabetes mellitus with neuro comps	Probable T2 codes
13279	C104y00	Other specified diabetes mellitus with renal complications	Vague codes
45919	C109212	Type 2 diabetes mellitus with neurological complications	Probable T2 codes
47649	C10E100	Type 1 diabetes mellitus with ophthalmic complications	Definite T1 codes
49276	C108100	Insulin-dependent diabetes mellitus with ophthalmic comps	Probable T1 codes
53630	C110.11	Insulin coma	Possible T1 codes
65704	C109412	Type 2 diabetes mellitus with ulcer	Probable T2 codes
69278	C109E00	Non-insulin depend diabetes mellitus with diabetic cataract	Probable T2 codes
55431	L180X00	Pre-existing diabetes mellitus, unspecified	Possible T2 codes
41716	C108C00	Insulin dependent diabetes mellitus with polyneuropathy	Probable T1 codes
45276	C10E312	Insulin dependent diabetes mellitus with multiple complicat	Definite T1 codes
45467	C109B00	Non-insulin dependent diabetes mellitus with polyneuropathy	Probable T2 codes
52104	C108300	Insulin dependent diabetes mellitus with multiple complicatn	Probable T1 codes
70821	C10yz00	Diabetes mellitus NOS with other specified manifestation	Vague codes
105434	C11y400		Vague codes
24836	C109C12	Type 2 diabetes mellitus with nephropathy	Probable T2 codes
38161	C108711	Type I diabetes mellitus with retinopathy	Probable T1 codes
48192	C109E11	Type II diabetes mellitus with diabetic cataract	Probable T2 codes

51957	C108511	Type I diabetes mellitus with ulcer	Probable T1 codes
60499	C108600	Insulin dependent diabetes mellitus with gangrene	Probable T1 codes
64283	C10zy00	Other specified diabetes mellitus with unspecified comps	Vague codes
66965	C109H12	Type 2 diabetes mellitus with neuropathic arthropathy	Probable T2 codes
95351	C10FA11	Type II diabetes mellitus with mononeuropathy	Definite T2 codes
18642	C10EH00	Type 1 diabetes mellitus with arthropathy	Definite T1 codes
40401	C109500	Non-insulin dependent diabetes mellitus with gangrene	Probable T2 codes
40962	C109H00	Non-insulin dependent d m with neuropathic arthropathy	Probable T2 codes
41049	C108712	Type 1 diabetes mellitus with retinopathy	Probable T1 codes
43227	C10F311	Type II diabetes mellitus with multiple complications	Definite T2 codes
43785	C109D00	Non-insulin dependent diabetes mellitus with hypoglyca coma	Probable T2 codes
46290	C108y00	Other specified diabetes mellitus with multiple comps	Vague codes
50225	C109011	Type II diabetes mellitus with renal complications	Probable T2 codes
62146	C109300	Non-insulin-dependent diabetes mellitus with multiple comps	Probable T2 codes
67635	L180000	Diabetes mellitus - unspec whether in pregnancy/puerperium	Probable Gestational diabetes
93727	C10FE11	Type II diabetes mellitus with diabetic cataract	Definite T2 codes
94383	C10N000	Secondary diabetes mellitus without complication	Secondary / Other types
96235	C10E911	Type I diabetes mellitus maturity onset	Definite T1 codes
10098	C10yy00	Other specified diabetes mellitus with other spec comps	Vague codes
52283	C108200	Insulin-dependent diabetes mellitus with neurological comps	Probable T1 codes
69124	C107300	IDDM with peripheral circulatory disorder	Probable T1 codes
69993	C10E600	Type 1 diabetes mellitus with gangrene	Definite T1 codes
18209	C109012	Type 2 diabetes mellitus with renal complications	Probable T2 codes
40023	C102000	Diabetes mellitus, juvenile type, with hyperosmolar coma	Possible T1 codes
44260	C108F00	Insulin dependent diabetes mellitus with diabetic cataract	Probable T1 codes
44779	C109E12	Type 2 diabetes mellitus with diabetic cataract	Probable T2 codes
47816	C109H11	Type II diabetes mellitus with neuropathic arthropathy	Probable T2 codes
62107	C109511	Type II diabetes mellitus with gangrene	Probable T2 codes
63371	C10y100	Diabetes mellitus, adult, + other specified manifestation	Probable T2 codes
68792	C10z000	Diabetes mellitus, juvenile type, + unspecified complication	Possible T1 codes
72345	C102z00	Diabetes mellitus NOS with hyperosmolar coma	Vague codes
98723	C10FD11	Type II diabetes mellitus with hypoglycaemic coma	Definite T2 codes
102201	C10FC11	Type II diabetes mellitus with nephropathy	Definite T2 codes
42567	C103000	Diabetes mellitus, juvenile type, with ketoacidotic coma	Possible T1 codes
54212	C109F00	Non-insulin-dependent d m with peripheral angiopath	Probable T2 codes
61071	C109D12	Type 2 diabetes mellitus with hypoglycaemic coma	Probable T2 codes
65062	C103z00	Diabetes mellitus NOS with ketoacidotic coma	Vague codes
68105	C10EB00	Type 1 diabetes mellitus with mononeuropathy	Definite T1 codes
69748	C105000	Diabetes mellitus, juvenile type, + ophthalmic manifestation	Possible T1 codes
93878	C10E511	Type I diabetes mellitus with ulcer	Definite T1 codes
97849	C10E912	Insulin dependent diabetes maturity onset	Definite T1 codes
21983	C108012	Type 1 diabetes mellitus with renal complications	Probable T1 codes
24693	C109G00	Non-insulin dependent diabetes mellitus with arthropathy	Probable T2 codes
46150	C109512	Type 2 diabetes mellitus with gangrene	Probable T2 codes
47377	C105y00	Other specified diabetes mellitus with ophthalmic complicatn	Vague codes
49949	C10E411	Unstable type I diabetes mellitus	Definite T1 codes
61344	C108011	Type I diabetes mellitus with renal complications	Probable T1 codes
64571	C109C11	Type II diabetes mellitus with nephropathy	Probable T2 codes
72702	C10E812	Insulin dependent diabetes mellitus - poor control	Definite T1 codes
91646	C10F411	Type II diabetes mellitus with ulcer	Definite T2 codes
93875	C10E712	Insulin dependent diabetes mellitus with retinopathy	Definite T1 codes
95539	C10FS00	Maternally inherited diabetes mellitus	Genetic
98392	C10C.12	Maturity onset diabetes in youth type 1	Genetic
102112	C10E611	Type I diabetes mellitus with gangrene	Definite T1 codes
103902	C10FG11	Type II diabetes mellitus with arthropathy	Definite T2 codes
17545	C108F11	Type I diabetes mellitus with diabetic cataract	Probable T1 codes
50527	C10FB11	Type II diabetes mellitus with polyneuropathy	Definite T2 codes
54600	C10E412	Unstable insulin dependent diabetes mellitus	Definite T1 codes
54899	C109F11	Type II diabetes mellitus with peripheral angiopathy	Probable T2 codes
57278	C10F011	Type II diabetes mellitus with renal complications	Definite T2 codes
61829	C108212	Type 1 diabetes mellitus with neurological complications	Probable T1 codes
66872	C108D11	Type I diabetes mellitus with nephropathy	Probable T1 codes
72882	C110100	Self-induced hyperinsulinism	Possible T2 codes
85991	C10FM11	Type II diabetes mellitus with persistent microalbuminuria	Definite T2 codes
95343	C10E711	Type I diabetes mellitus with retinopathy	Definite T1 codes
18230	C108J12	Type 1 diabetes mellitus with neuropathic arthropathy	Probable T1 codes
47409	C109B11	Type II diabetes mellitus with polyneuropathy	Probable T2 codes

59288	C103y00	Other specified diabetes mellitus with coma	Vague codes
60107	C108411	Unstable type I diabetes mellitus	Probable T1 codes
60208	C108J11	Type I diabetes mellitus with neuropathic arthropathy	Probable T1 codes
64446	C108G00	Insulin dependent diab mell with peripheral angiopathy	Probable T1 codes
64449	C108z00	Unspecified diabetes mellitus with multiple complications	Vague codes
65616	C108H00	Insulin dependent diabetes mellitus with arthropathy	Probable T1 codes
66675	C10A000	Malnutrition-related diabetes mellitus with coma	Secondary / Other types
67905	C109211	Type II diabetes mellitus with neurological complications	Probable T2 codes
68517	C10J.00	Insulin autoimmune syndrome	Secondary / Other types
91942	C10E311	Type I diabetes mellitus with multiple complications	Definite T1 codes
100964	C10F111	Type II diabetes mellitus with ophthalmic complications	Definite T2 codes
45914	C108812	Type 1 diabetes mellitus - poor control	Probable T1 codes
56268	C109D11	Type II diabetes mellitus with hypoglycaemic coma	Probable T2 codes
59725	C109111	Type II diabetes mellitus with ophthalmic complications	Probable T2 codes
60699	C109F12	Type 2 diabetes mellitus with peripheral angiopathy	Probable T2 codes
62613	C10EA11	Type I diabetes mellitus without complication	Definite T1 codes
67212	C10H000	DM induced by non-steroid drugs without complication	Secondary / Other types
68843	C103100	Diabetes mellitus, adult onset, with ketoacidotic coma	Possible T2 codes
69908	C11yz00	Oth spec int pancr secret NOS	Secondary / Other types
70448	C107000	Diabetes mellitus, juvenile +peripheral circulatory disorder	Possible T1 codes
70766	C108E12	Type 1 diabetes mellitus with hypoglycaemic coma	Probable T1 codes
72320	C109A00	Non-insulin dependent diabetes mellitus with mononeuropathy	Probable T2 codes
93468	C10EG00	Type 1 diabetes mellitus with peripheral angiopathy	Definite T1 codes
93922	C104000	Diabetes mellitus, juvenile type, with renal manifestation	Possible T1 codes
96823	L180400	Diabetes mellitus in pueperium - baby previously delivered	Probable Gestational diabetes
97446	C108912	Type 1 diabetes mellitus maturity onset	Probable T1 codes
97474	C108412	Unstable type 1 diabetes mellitus	Probable T1 codes
98616	C10F211	Type II diabetes mellitus with neurological complications	Definite T2 codes
99719	C10EA12	Insulin-dependent diabetes without complication	Definite T1 codes
102946	C10E012	Insulin-dependent diabetes mellitus with renal complications	Definite T1 codes
105337	C10E811	Type I diabetes mellitus - poor control	Definite T1 codes
18143	C109G11	Type II diabetes mellitus with arthropathy	Probable T2 codes
24694	C108B00	Insulin dependent diabetes mellitus with mononeuropathy	Probable T1 codes
49146	C108211	Type I diabetes mellitus with neurological complications	Probable T1 codes
49869	C109G12	Type 2 diabetes mellitus with arthropathy	Probable T2 codes
50813	C109A11	Type II diabetes mellitus with mononeuropathy	Probable T2 codes
56885	C10K000	Type A insulin resistance without complication	Not diabetes
62352	C108H11	Type I diabetes mellitus with arthropathy	Probable T1 codes
63017	C108911	Type I diabetes mellitus maturity onset	Probable T1 codes
66145	C10EN11	Type I diabetes mellitus with ketoacidotic coma	Definite T1 codes
70316	C109112	Type 2 diabetes mellitus with ophthalmic complications	Probable T2 codes
91943	C10EC11	Type I diabetes mellitus with polyneuropathy	Definite T1 codes
95992	C108A11	Type I diabetes mellitus without complication	Probable T1 codes
96506	C10G000	Secondary pancreatic diabetes mellitus without complication	Secondary / Other types
97894	C10EP11	Type I diabetes mellitus with exudative maculopathy	Definite T1 codes
98071	C10E112	Insulin-dependent diabetes mellitus with ophthalmic comps	Definite T1 codes
98704	C10E512	Insulin dependent diabetes mellitus with ulcer	Definite T1 codes
99231	C108B11	Type I diabetes mellitus with mononeuropathy	Probable T1 codes
99311	C10E111	Type I diabetes mellitus with ophthalmic complications	Definite T1 codes
99716	C10EE12	Insulin dependent diabetes mellitus with hypoglycaemic coma	Definite T1 codes
100347	C10A500	Malnutritn-relat diabetes melitus wth periph circul complctn	Secondary / Other types
100770	C10EF12	Insulin dependent diabetes mellitus with diabetic cataract	Definite T1 codes
101311	C10EC12	Insulin dependent diabetes mellitus with polyneuropathy	Definite T1 codes
101735	C10E212	Insulin-dependent diabetes mellitus with neurological comps	Definite T1 codes
102163	C10ED12	Insulin dependent diabetes mellitus with nephropathy	Definite T1 codes
102620	C10EL11	Type I diabetes mellitus with persistent microalbuminuria	Definite T1 codes
102740	C108112	Type 1 diabetes mellitus with ophthalmic complications	Probable T1 codes
104639	C10FF11	Type II diabetes mellitus with peripheral angiopathy	Definite T2 codes
75579	L1801G		Probable Gestational diabetes
77675	L1801T		Probable Gestational diabetes
82025	L1801NM		Probable Gestational diabetes
104323	C10F511	Type II diabetes mellitus with gangrene	Definite T2 codes
1684	66A3.00	Diabetic on diet only	Probable T2 codes
7563	66A4.00	Diabetic on oral treatment	Probable T2 codes
8842	66A5.00	Diabetic on insulin	Possible T1 codes

Process of care codes

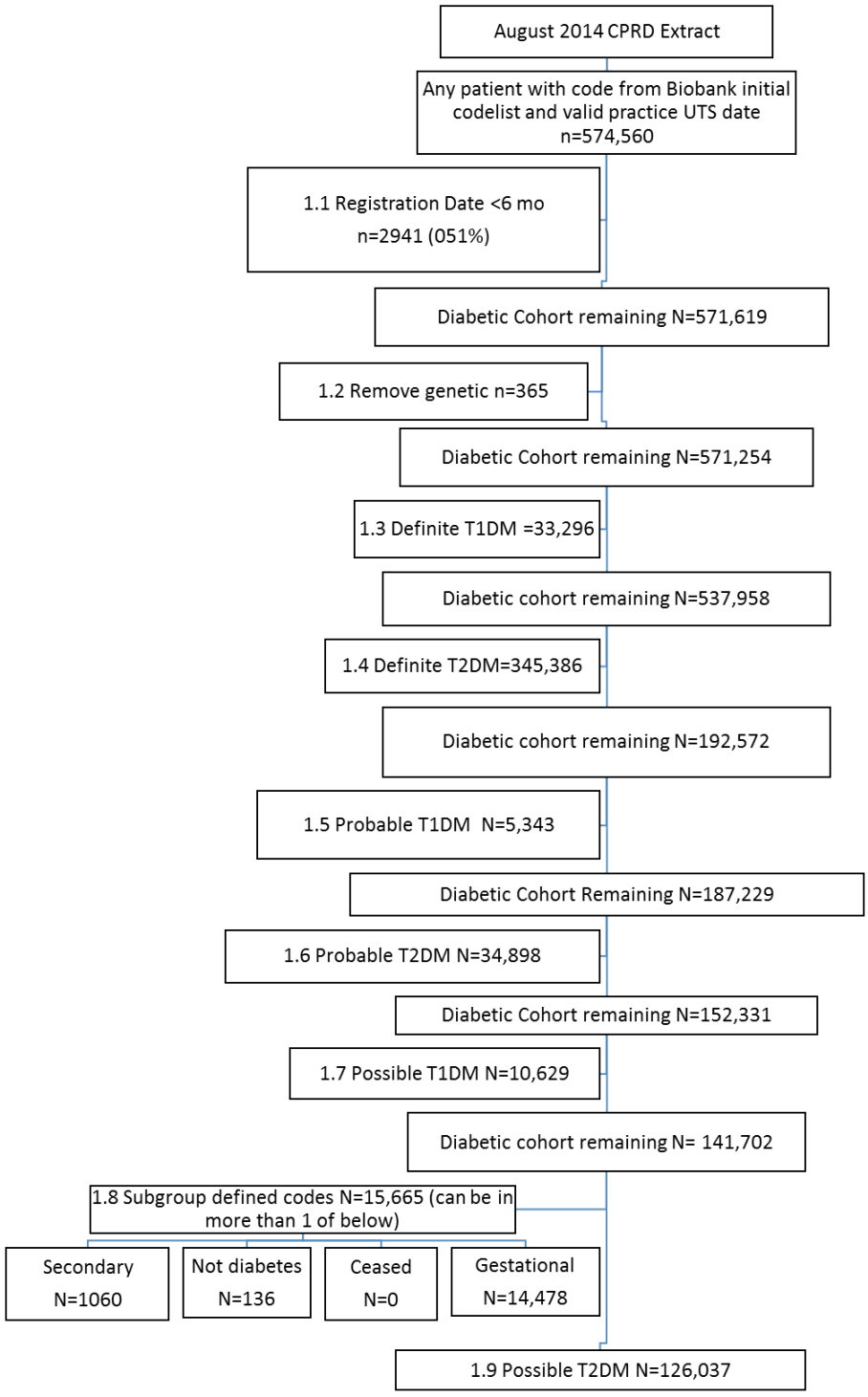
readcode	readterm	readcode	readterm
1434	H/O: diabetes mellitus	**585Y.00**	Right posterior tibial doppler pressure
14F4.00	H/O: Admission in last year for diabetes foot problem	**585a.00**	ABPI - Ankle brachial pressure index
14P3.00	H/O: insulin therapy	**585b.00**	Left dorsalis pedis ABPI
24E..00	O/E - peripheral pulses R.-leg	**585c.00**	Right dorsalis pedis ABPI
24E..11	O/E - dorsalis pedis -R	**585d.00**	Left posterior tibial ABPI
24E..12	O/E - femoral pulse - R	**585e.00**	Right posterior tibial ABPI
24E..13	O/E - popliteal pulse-R	**58C1.00**	Retinal photography
24E..14	O/E - post-tibial pulse - R	**66A..00**	Diabetic monitoring
24E..15	O/E - pulses - R-leg	**66A1.00**	Initial diabetic assessment
24E1.00	O/E -R.-leg pulses all present	**66A2.00**	Follow-up diabetic assessment
24E2.00	O/E - R.femoral pulse present	**66A6.00**	Last hypo. attack
24E3.00	O/E - R.femoral pulse absent	**66A7.00**	Frequency of hypo. attacks
24E5.00	O/E - R.popliteal pulse absent	**66A7000**	Frequency of hospital treated hypoglycaemia
24E6.00	O/E - R.post.tib.pulse present	**66A7100**	Frequency of GP or paramedic treated hypoglycaemia
24E7.00	O/E - R.post.tib pulse absent	**66A8.00**	Has seen dietician - diabetes
24E8.00	O/E - R.dorsalis pedis present	**66A9.00**	Understands diet - diabetes
24E9.00	O/E - R.dorsalis pedis absent	**66AA.11**	Injection sites - diabetic
24EA.00	O/E - Absent right foot pulses	**66AB.00**	Urine sugar charts
24EB.00	O/E - right foot pulses present	**66AC.00**	Blood sugar charts
24EC.00	O/E - Right dorsalis pedis abnormal	**66AD.00**	Fundoscopy - diabetic check
24ED.00	O/E - Right posterior tibial pulse abnormal	**66AE.00**	Feet examination
24EE.00	O/E - Right dorsalis pedis normal	**66AF.00**	Attends out-patients
24EF.00	O/E - Right posterior tibial pulse normal	**66AG.00**	Diabetic drug side effects
24EZ.00	O/E - R.leg pulses NOS	**66AH.00**	Diabetic treatment changed
24F1.00	O/E - L.leg pulses all present	**66AH000**	Conversion to insulin
24F2.00	O/E - L.femoral pulse present	**66AI.00**	Diabetic - good control
24F3.00	O/E - L.femoral pulse absent	**66AJ.00**	Diabetic - poor control
24F4.00	O/E -L.popliteal pulse present	**66AJ.11**	Unstable diabetes
24F5.00	O/E - L.popliteal pulse absent	**66AJ000**	Chronic hyperglycaemia
24F6.00	O/E - L.post.tib.pulse present	**66AJ100**	Brittle diabetes
24F7.00	O/E - L.post.tib. pulse absent	**66AJ200**	Loss of hypoglycaemic warning
24F8.00	O/E - L.dorsalis pedis present	**66AJ300**	Recurrent severe hypos
24F9.00	O/E - L.dorsalis pedis absent	**66AJz00**	Diabetic - poor control NOS
24FA.00	O/E - Absent left foot pulses	**66AK.00**	Diabetic - cooperative patient
24FB.00	O/E - left foot pulses present	**66AL.00**	Diabetic-uncooperative patient
24FC.00	O/E - left dorsalis pedis abnormal	**66AM.00**	Diabetic - follow-up default
24FD.00	O/E - Left posterior tibial pulse abnormal	**66AN.00**	Date diabetic treatment start
24FE.00	O/E - left dorsalis pedis normal	**66AO.00**	Date diabetic treatment stopp.
24FF.00	O/E - Left posterior tibial pulse normal	**66AP.00**	Diabetes: practice programme
29B1.00	O/E - tactile sensation normal	**66AQ.00**	Diabetes: shared care programme
29B2.00	O/E - anaesthesia present	**66AQ000**	Unsuitable for diabetes year of care programme
29B2.11	O/E - loss of touch sensation	**66AQ100**	Declined consent for diabetes year of care programme
29B2000	O/E - anaesthesia in legs	**66AR.00**	Diabetes management plan given
29B2100	O/E - anaesthesia of extremities	**66AS.00**	Diabetic annual review
29B3.00	O/E - hypoaesthesia present	**66AT.00**	Annual diabetic blood test
29B4.00	O/E - hyperaesthesia present	**66AU.00**	Diabetes care by hospital only
29B4.11	O/E - hyperalgesia present	**66AV.00**	Diabetic on insulin and oral treatment
29B5.00	O/E - paraesthesia present	**66AW.00**	Diabetic foot risk assessment
29B5000	O/E - paraesthesia in hands	**66AX.00**	Diabetes: shared care in pregnancy - diabetol and obstet
29B6.00	Hemisensory loss	**66AY.00**	Diabetic diet - good compliance
29B7.00	10g monofilament sensation present	**66AZ.00**	Diabetic monitoring NOS
29B8.00	10g monofilament sensation absent	**66Aa.00**	Diabetic diet - poor compliance
29B9.00	10g monofilament sensation R foot abnormal	**66Ab.00**	Diabetic foot examination
29BA.00	10g monofilament sensation L foot abnormal	**66Ac.00**	Diabetic peripheral neuropathy screening
29BB.00	10g monofilament sensation R foot normal	**66Ad.00**	Hypoglycaemic attack requiring 3rd party assistance
29BC.00	10g monofilament sensation L foot normal	**66Ae.00**	HbA1c target
29BD.00	10g monofil sensation plantar aspect gt toe L foot present	**66Ae000**	HbA1c target level - IFCC standardised

29BE.00	10g monofil sensation plantar aspect mid toe R foot present	**66Af.00**	Patient diabetes education review
29BF.00	10g monofil sensation plantar aspect mid toe L foot present	**66Ag.00**	Insulin needles changed daily
29BG.00	10g monofil sensation plantar aspect lit toe R foot present	**66Ah.00**	Insulin needles changed for each injection
29BM.00	10g monofil sensation plantar aspect gt toe R foot absent	**66Ai.00**	Diabetic 6 month review
29BN.00	10g monofil sensation plantar aspect gt toe L foot absent	**66Aj.00**	Insulin needles changed less than once a day
29BQ.00	10g monofil sensation plantar aspect mid toe L foot absent	**66Ak.00**	Diabetic monitoring - lower risk albumin excretion
29H1.00	O/E - vibration sense normal	**66Al.00**	Diabetic monitoring - higher risk albumin excretion
29H2.00	O/E - vibration sense reduced	**66Am.00**	Insulin dose changed
29H3.00	O/E - vibration sense absent	**66An.00**	Diabetes type 1 review
29H4.00	O/E - Vibration sense of right foot abnormal	**66Ao.00**	Diabetes type 2 review
29H5.00	O/E - Vibration sense of right foot normal	**66Ap.00**	Insulin treatment initiated
29H6.00	O/E - Vibration sense of left foot abnormal	**66Aq.00**	Diabetic foot screen
29H7.00	O/E - Vibration sense of left foot normal	**66Ar.00**	Insulin treatment stopped
29H8.00	O/E - vibration sense left foot reduced	**66As.00**	Diabetic on subcutaneous treatment
29H9.00	O/E - vibration sense right foot reduced	**66At.00**	Diabetic dietary review
29HA.00	O/E - Vibration sense of right foot absent	**66At000**	Type I diabetic dietary review
29HB.00	O/E - Vibration sense of left foot absent	**66At011**	Type 1 diabetic dietary review
2BB..00	O/E - retinal inspection	**66At100**	Type II diabetic dietary review
2BB..11	O/E - retina	**66At111**	Type 2 diabetic dietary review
2BB1.00	O/E - retina normal	**66Au.00**	Diabetic erectile dysfunction review
2BB2.00	O/E - retinal vessel narrowing	**66Av.00**	Diabetic assessment of erectile dysfunction
2BB3.00	O/E - retinal A-V nipping	**66Aw.00**	Insulin dose
2BB4.00	O/E - retinal microaneurysms	**66Ax.00**	Checking accuracy of blood glucose meter
2BB5.00	O/E - retinal haemorrhages	**66Ay.00**	Gestational diabetes mellitus annual review
2BB6.00	O/E - retinal exudates	**66Az.00**	High risk of diabetes mellitus annual review
2BB7.00	O/E - retinal vascular prolif.	**671F000**	Insulin alert pat information booklet information discussed
2BB8.00	O/E - vitreous haemorrhages	**679L.00**	Health education - diabetes
2BB9.00	O/E - retinal pigmentation	**679L000**	Education in self management of diabetes
2BBA.00	Examination of retina	**679L100**	
2BBB.00	O/E - Right retina not seen	**679c.00**	Insulin administration education
2BBC.00	O/E - Left retina not seen	**67IJ100**	Pre-conception advice for diabetes mellitus
2BBD.00	O/E - Right retina normal	**68A8.00**	Digital retinal screening
2BBE.00	O/E - Left retina normal	**68A9.00**	Diabetic retinopathy screening offered
2BBF.00	Retinal abnormality - diabetes related	**68AA.00**	Digital retinal screening offered
2BBG.00	Retinal abnormality - non-diabetes	**68AB.00**	Diabetic digital retinopathy screening offered
2BBH.00	Retinal drusen	**7L10000**	Continuous subcutaneous infusion of insulin
2BBI.00	O/E - no retinopathy	**7L10011**	Subcutaneous infusion with insulin pump
2BBJ.00	O/E - no right diabetic retinopathy	**7L19800**	Subcutaneous injection of insulin
2BBK.00	O/E - no left diabetic retinopathy	**8B3l.00**	Diabetes medication review
2BBL.00	O/E - diabetic maculopathy present both eyes	**8BAi.00**	Insulin passport completed
2BBM.00	O/E - diabetic maculopathy absent both eyes	**8BAm.00**	Insulin passport checked
2BBN.00	Myelinated retinal nerve fibres	**8BL2.00**	Patient on maximal tolerated therapy for diabetes
2BBO.00	O/E - Laser photocoagulation scars	**8CA4100**	Pt advised re diabetic diet
2BBP.00	O/E - right eye background diabetic retinopathy	**8CE0100**	Insulin alert patient information booklet given
2BBQ.00	O/E - left eye background diabetic retinopathy	**8CE0200**	Insulin passport given
2BBR.00	O/E - right eye preproliferative diabetic retinopathy	**8CMW700**	Diabetes clinical pathway
2BBS.00	O/E - left eye preproliferative diabetic retinopathy	**8CS0.00**	Diabetes care plan agreed
2BBT.00	O/E - right eye proliferative diabetic retinopathy	**8H7f.00**	Referral to diabetes nurse
2BBV.00	O/E - left eye proliferative diabetic retinopathy	**8H7r.00**	Refer to diabetic foot screener
2BBW.00	O/E - right eye diabetic maculopathy	**8HBD.00**	Retinopathy follow up
2BBX.00	O/E - left eye diabetic maculopathy	**8HBG.00**	Diabetic retinopathy 12 month review
2BBY.00	O/E - referable retinopathy	**8HBH.00**	Diabetic retinopathy 6 month review
2BBZ.00	O/E - retinal inspection NOS	**8HgC.00**	Discharged from diabetes shared care programme
2BBa.00	O/E- non-referable retinopathy	**8Hj0.00**	Referral to diabetes structured education programme
2BBb.00	O/E - fundus not adequately seen	**8Hl4.00**	Referral to community diabetes specialist nurse
2BBc.00	O/E - No retinal laser photocoagulation scars	**8Hlc.00**	Referral to community diabetes service
2BBd.00	O/E - Red reflex absent	**8I3W.00**	Diabetic foot examination declined
2BBe.00	O/E - right retina partially assessable	**8I3X.00**	Diabetic retinopathy screening refused

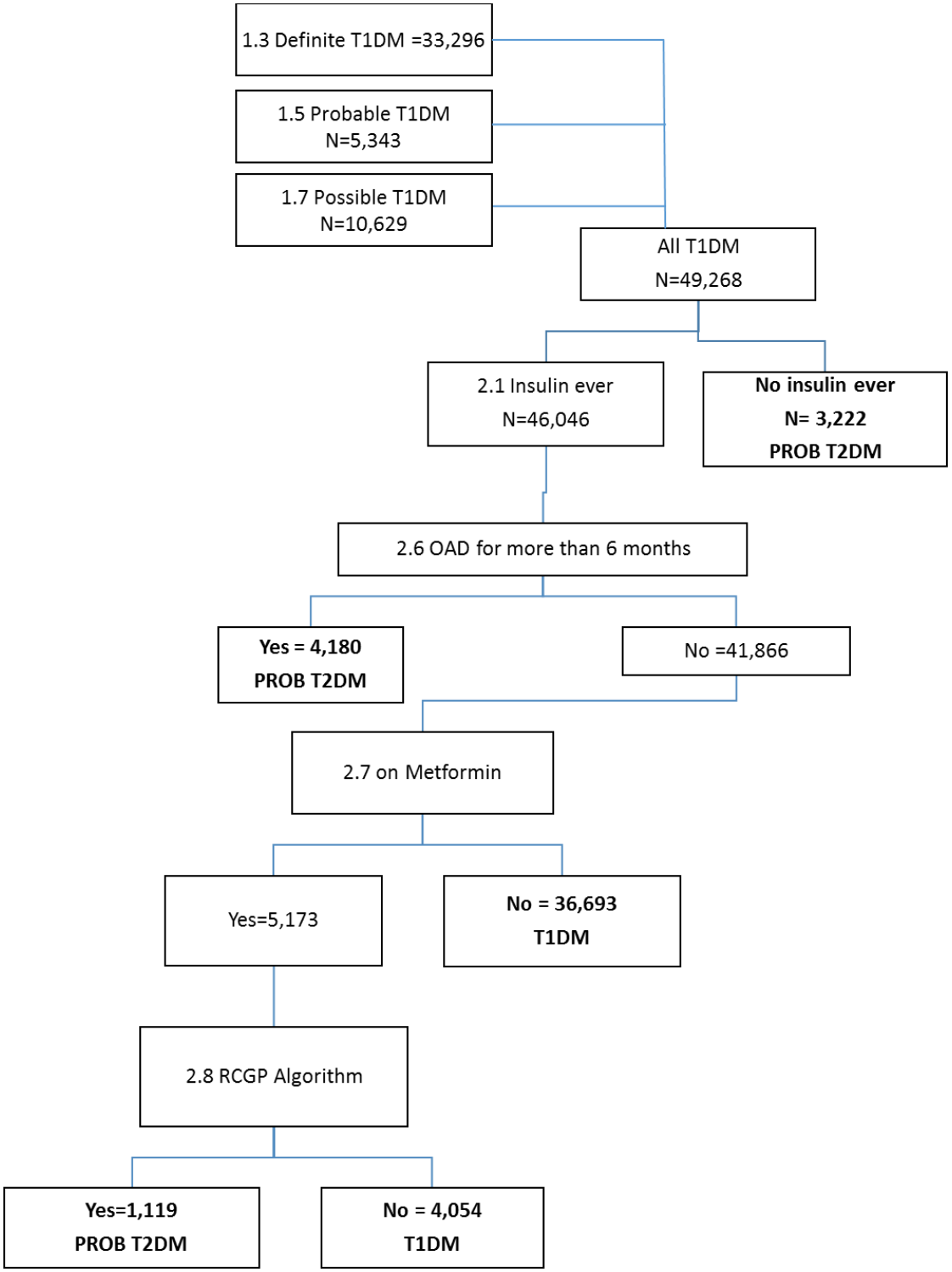
2BBf.00	O/E - left retina partially assessable	**8I3k.00**	Insulin therapy declined
2BBg.00	O/E - right retina fully assessable	**8I6F.00**	Diabetic retinopathy screening not indicated
2BBh.00	O/E - left retina fully assessable	**8I6G.00**	Diabetic foot examination not indicated
2BBi.00	O/E - right eye no maculopathy	**8IF..00**	Prof judgemnt not to engage pt wt insulin alert requirements
2BBj.00	O/E - left eye no maculopathy	**9M00.00**	Informed consent for diabetes national audit
2BBk.00	O/E - right eye stable treated prolif diabetic retinopathy	**9M10.00**	Informed dissent for diabetes national audit
2BBl.00	O/E - left eye stable treated prolif diabetic retinopathy	**9N0o.00**	Seen in community diabetic specialist nurse clinic
2BBm.00	O/E - right eye clinically significant macular oedema	**9N1v.00**	Seen in diabetic eye clinic
2BBn.00	O/E - left eye clinically significant macular oedema	**9N2f.00**	Seen by retinal screener
2BBo.00	O/E - sight threatening diabetic retinopathy	**9NM0.00**	Attending diabetes clinic
2BBp.00	On examination right red reflex present	**9NNC.00**	Under care of retinal screener
2BBq.00	On examination left red reflex present	**9NND.00**	Under care of diabetic foot screener
2BBr.00	Impaired vision due to diabetic retinopathy	**9OL..00**	Diabetes monitoring admin.
2BBs.00	Retinal arteries silverwire	**9OLD.00**	Diabetic patient unsuitable for digital retinal photography
311A.00	Monofilament foot sensation test	**9Oy0.00**	Diabetes screening invitation
3128	Fundoscopy	**9Oy0000**	
3128.11	Retinoscopy	**9Oy0200**	
3128000	Fundoscopy normal	**9Oy0300**	
3128100	Fundoscopy abnormal	**9kL..00**	**Insulin initiation - enhanced services administration**
3128200	Dilated fundoscopy normal	**9m00.00**	Eligible for diabetic retinopathy screening
3128300	Camera fundoscopy	**9m0A.00**	Declined diabetic retinopathy screening
3128400	Indirect fundoscopy following mydriatic	**M21yC00**	Insulin lipohypertrophy
3128Z00	Fundoscopy NOS	**M21yC11**	Insulin site lipohypertrophy
3129	Eye fundus photography	**U602311**	[X] Adverse reaction to insulins and antidiabetic agents
43Gk.00	Insulin antibody level	**U602312**	[X] Adverse reaction to insulins
449G.00	Serum pro-insulin level	**U60231E**	[X] Adverse reaction to insulins and antidiabetic agents NOS
585V.00	Left dorsalis pedis doppler pressure	**ZV6DA00**	[V]Admitted for commencement of insulin
585W.00	Right dorsalis pedis doppler pressure	**ZV6DB00**	[V]Admitted for conversion to insulin
585X.00	Left posterior tibial doppler pressure	**U60231E**	[X] Adverse reaction to insulins and antidiabetic agents NOS
ZV6DB00	[V]Admitted for conversion to insulin	**ZV6DA00**	[V]Admitted for commencement of insulin

APPENDIX 8 DETAILED BIOBANK ALGORITHM FLOWCHARTS

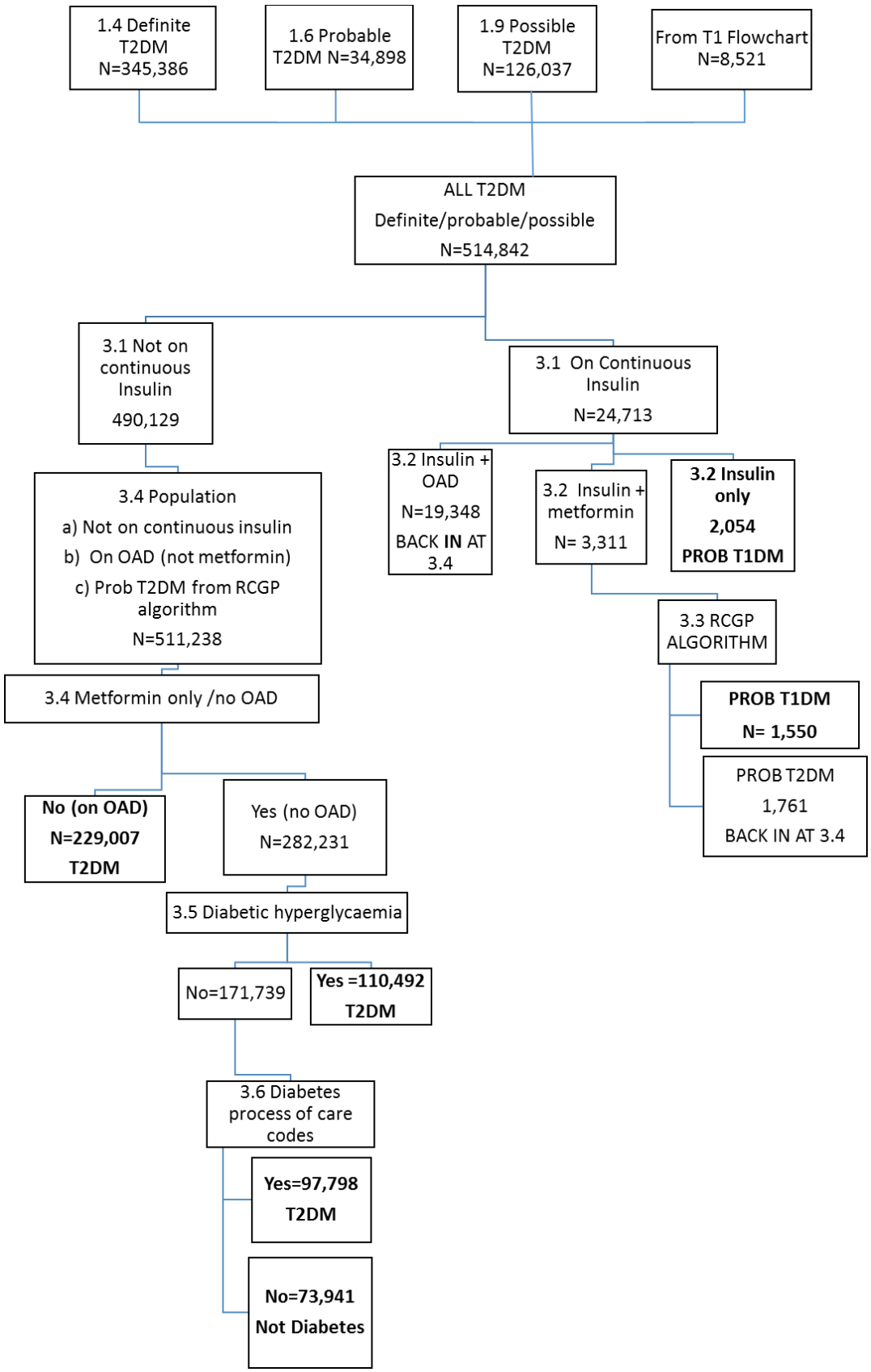
Stage one initial sorting process



Stage two Type one diabetes sort



Stage twoType 2 diabetes sort



APPENDIX 9 HOW ADDITIONAL INFORMATION NEEDED FOR BIOBANK ALGORITHM WAS EXTRACTED

Anti-diabetes medications:

Anti-diabetes medications were identified from the complete product code list using British National Formulary (BNF) codes. Specifically all codes starting 060101, 060102, and excluding 06010103 were extracted – relating to subchapters Insulin, Antidiabetic drugs, and excluding Hypodermic equipment respectively. Partial string matches were then used on the "drugsubstance" field to identify the different anti diabetic drug classes. They were also classified broadly as either biguanides (metformin), Insulin, or other oral anti diabetic agents (OADs). These product code lists were then merged with the therapy files to obtain longitudinal prescription histories of anti-diabetic medications. From this, it was possible to calculate for each subject, the start and end date (if any) each class of medication, and also indicators for any gaps in medication.

The exact code used to define the medication categories is detailed below in box 2.

```
gen diabetestreat = 0
replace diabetestreat=1 if strmatch(bnfcode, "060101*") ///
| strmatch(bnfcode, "060102*")
replace diabetestreat=0 if strmatch(bnfcode, "06010103*")
gen nodrug = 0
replace nodrug=1 if strmatch(drugsubstance ,"*None known*")
gen diab_therapy=.
replace productname = lower(productname)
replace bnfchapter = lower(bnfchapter)


replace diab_therapy=1 if regexm(bnfcode, "60101") |
regexm(productname, "insulin") | regexm(productname, "Insulin")
replace diab_therapy=2 if regexm(bnfcode, "60102")
replace diab_therapy=3 if ( strmatch(drugsubstance, "*metformin*")
| regexm(productname, "Metformin") ) & strmatch(drugsubstance,
"*/*")!=1
```

Other OADs (diab_therapy=2), were then further classified using partial string matches and bnfcodes as detailed below, and checked manually.

Sulfonylureas: bnfcode 6.1.2.1 (plus partial string match for drug substance with “Gli” or ending in “mide”)

Glitazones: string search for drug substance ending in “zone”

Glinides: string search for drug substance ending in “glinide”

DPP4: string search for drug sunstance ending in “gliptin”

GLPs: string search for drug substance ending in “tide” excluding pramlintide

Anything else was categorised as “other”.

BMI

Height and weight were extracted by searching the additional information file for entity types 13 and 14. These were then merged with the clinical information file to get dates of measurements.

Height was cleaned by first searching for values >100, suggesting height had been entered in cm. These were converted to m. After this, entries suggesting implausible values (>2.3 m (7.5 feet) and <1.3 m (4 feet)) were set to missing. Multiple height entries on the same day were replaced with the mean if the difference was less than 5 cm, or both created missing.

Weight was then cleaned in a similar way, with implausible values removed (<4 stone (25kg) or > 40 stone (255kg)), and multiple entries on the same day replaced with means, or removed if the differences between the two results differed by 2kg or more.

BMI was then calculated using the cleaned height and weight measures, and the value closest to but preceding the diagnosis code was selected as the BMI used in the confirmation algorithm.

Hyperglycaemia

HbA1c and glucose test results were extracted from the test file using entity types 275 and 274 respectively.

For HbA1c, results were only kept if they were entered as %, mmol/L, iu/L and mmol/mol. (this accounted for 80% of all entries, with a further 19% being labelled as “unit not entered”). All results were then converted to %. After conversion, readings of <2% and >20 % were considered implausible. Duplicates on the same day were then identified. Any result originally in % format was given priority. Multiple readings in % were then compared and removed if the difference was > 2%. Otherwise a mean was used.

Fasting glucose values were all entered in the same unit (mmol/L). Duplicates by patient and date were removed using a similar method, with a difference of 2 mmol/L as the maximum difference allowed.

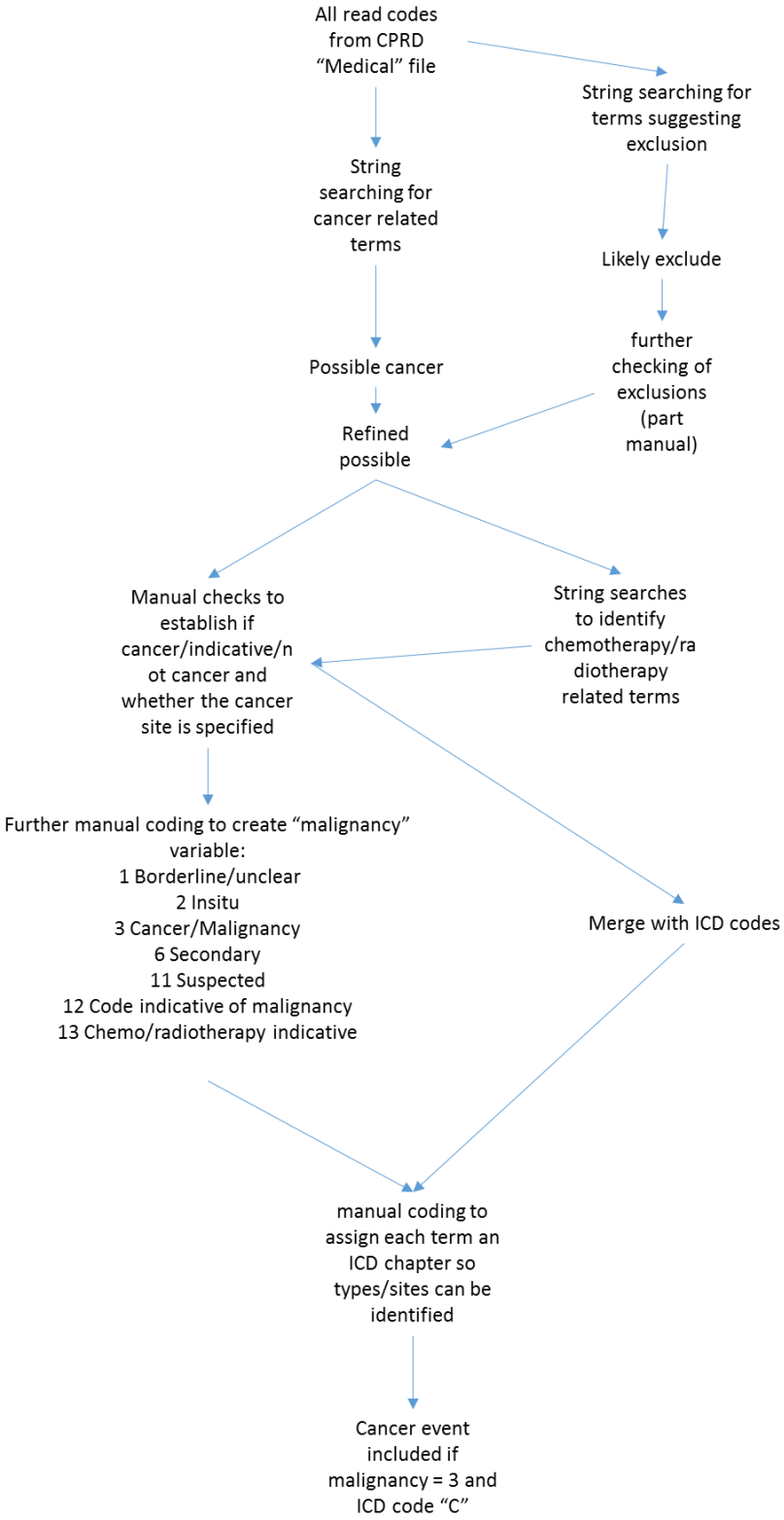
The closest HbA1c and glucose value to the diagnosis code were taken as those used to define whether the patient had hyperglycaemia, defined as an HbA1c > 7.5 or fasting glucose > 7.1. Longitudinal files of HbA1c were saved for use in later chapters.

APPENDIX 10 SUMMARY OF FIRST AND SECOND LINE TREATMENT OPTIONS FOR PATIENTS DIAGNOSED WITH T2DM AFTER JANUARY 2005.

	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	37725 (82.31)	9 , 1 , 1 – 9	1886 (10.5)	15.7 , 10 , 3 – 23	0 (0)		1,687 (83)	15.6 , 10 , 3 – 23
Sulfonylurea	3615 (7.89)	6.2 , 1 , 1 – 2	9558 (53.2)	20 , 14 , 5 – 30	9446 (65.8)	20.1 , 14 , 5 – 30		
Insulin	217 (0.47)	6.3 , 1 , 1 – 4	586 (3.3)	15.5 , 8 , 2 – 20	152 (1.1)	15.7 , 9 , 2 – 21	178 (8.8)	12.9 , 5.5 , 2 – 17
TZDs	83 (0.18)	10.8 , 1 , 1 – 13	1681 (9.4)	21.7 , 17 , 7 – 33	1371 (9.6)	21.8 , 17 , 6 – 33	61 (3)	18.9 , 12 , 4 – 31
Glinides	20 (0.04)	10.2 , 1 , 1 – 4.5	80 (0.4)	14.5 , 11 , 4 – 20.5	56 (0.4)	15.4 , 12 , 5.5 – 20.5	7 (0.3)	11.3 , 8 , 5 – 19
GLPs	9 (0.02)	32.2 , 16 , 5 – 53	332 (1.8)	24.1 , 20 , 8 – 34.5	223 (1.6)	24.2 , 21 , 8 – 36	1 (0)	9 , 9 , 9 – 9
DPP4	83 (0.18)	14.7 , 1 , 1 – 26	3076 (17.1)	27 , 22 , 9 – 40	2482 (17.3)	26.6 , 22 , 9 – 39	69 (3.4)	27.7 , 22 , 12 – 43
Other	28 (0.06)	13.8 , 1 , 1 – 12	559 (3.1)	27.7 , 23 , 9 – 40	474 (3.3)	27 , 22 , 9 – 40	9 (0.4)	27.4 , 24 , 14 – 37
Metformin/Sulf combination	3353 (7.32)	2.5 , 1 , 1 – 1	7 (0)	5.1 , 4 , 2 – 8				
Metformin/Insulin	116 (0.25)	1.8 , 1 , 1 – 1	6 (0)	6.8 , 2.5 , 1 – 5			6 (0.3)	6.8 , 2.5 , 1 – 5
Metformin/Other	269 (0.59)	4.6 , 1 , 1 – 1	17 (0.1)	14.8 , 9 , 4 – 24			14 (0.7)	17.2 , 17 , 6 – 27
Sulfonylurea/Insulin	112 (0.24)	2.2 , 1 , 1 – 1	33 (0.2)	12.8 , 7 , 1 – 15	33 (0.2)	12.8 , 7 , 1 – 15		
Sulfonylurea/Other	25 (0.05)	6.5 , 1 , 1 – 1	106 (0.6)	23 , 17 , 3 – 37	101 (0.7)	21.7 , 15 , 3 – 33		
Insulin/Other	2 (0)	10 , 10 , 1 – 19	17 (0.1)	24.9 , 12 , 2 – 33	3 (0)	29 , 10 , 1 – 76	1 (0)	2 , 2 , 2 – 2
Other dual combination	0 (0)		17 (0.1)	18.7 , 18 , 3 – 28	10 (0.1)	17.8 , 12.5 , 3 – 28	0 (0)	
3 or more	175 (0.38)	1.5 , 1 , 1 – 1	3 (0)	31 , 15 , 14 – 64	2 (0)	14.5 , 14.5 , 14 – 15	0 (0)	
Sub Total	**45,832 (100)**	**8.2 , 1 , 1 – 7**	**17,964 (100)**	**21.1 , 15 , 5 – 32**	**14,353 (100)**	**21.6 , 16 , 5 – 32**	**2,033**	**15.9 , 10 , 3 – 23**
None	20,654	43.7, 34,14-64	27,868	32.9 , 26, 11 – 49	23,372	32.9 , 27 , 12-49	1,582	31.7 , 23 , 9-48
Total	**66,486**		**45,832**		**37,725**		**3,615**	

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

APPENDIX 11 CANCER IDENTIFICATION ALGORITHM AS DEVELOPED BY BHASKARAN ET AL (2014) [194]



APPENDIX 12 CODE LISTS AND DETAILS OF HOW ALL COVARIATES DEFINED WHERE APPLICABLE

Cardiovascular disease

Code list for indicating history of or new onset of any CVD

medcode	readcode	readterm	medcode	readcode	readterm
240	G3...00	Ischaemic heart disease	**24783**	G3...11	Arteriosclerotic heart disease
241	G30..00	Acute myocardial infarction	**25842**	G33z.00	Angina pectoris NOS
504	G65..00	Transient cerebral ischaemia	**26863**	G33z600	New onset angina
1204	G30..14	Heart attack	**27951**	G31..00	Other acute and subacute ischaemic heart disease
1298	G66..11	CVA unspecified	**27977**	G31yz00	Other acute and subacute ischaemic heart disease NOS
1344	G340.12	Coronary artery disease	**28138**	G34..00	Other chronic ischaemic heart disease
1414	G33z300	Angina on effort	**28314**	G61X000	Left sided intracerebral haemorrhage, unspecified
1430	G33..00	Angina pectoris	**28554**	G33zz00	Angina pectoris NOS
1431	G311.13	Unstable angina	**28736**	G30y000	Acute atrial infarction
1433	G65..12	Transient ischaemic attack	**29421**	G344.00	Silent myocardial ischaemia
1469	G66..00	Stroke and cerebrovascular accident unspecified	**29643**	G303.00	Acute inferoposterior infarction
1655	G340.11	Triple vessel disease of the heart	**29758**	G30X.00	Acute transmural myocardial infarction of unspecif site
1676	G3z..00	Ischaemic heart disease NOS	**29902**	G330z00	Angina decubitus NOS
1677	G30..15	MI - acute myocardial infarction	**30045**	G616.00	External capsule haemorrhage
1678	G308.00	Inferior myocardial infarction NOS	**30330**	G309.00	Acute Q-wave infarct
1792	G3...13	IHD - Ischaemic heart disease	**30421**	G30..13	Cardiac rupture following myocardial infarction (MI)
1895	G65z.00	Transient cerebral ischaemia NOS	**31060**	G61X.00	Intracerebral haemorrhage in hemisphere, unspecified
2417	G65..13	Vertebro-basilar insufficiency	**31595**	G610.00	Cortical haemorrhage
2491	G30..12	Coronary thrombosis	**32272**	G38..00	Postoperative myocardial infarction
3132	G65..11	Drop attack	**32450**	G33z400	Ischaemic chest pain
3535	G61z.00	Intracerebral haemorrhage NOS	**32526**	14AA.00	H/O: heart disease NOS
3704	G307.00	Acute subendocardial infarction	**32854**	G30B.00	Acute posterolateral myocardial infarction
3999	G340000	Single coronary vessel disease	**33377**	G651.00	Vertebral artery syndrome
4017	G32..00	Old myocardial infarction	**33499**	G665.00	Pure motor lacunar syndrome
4656	G311.11	Crescendo angina	**33543**	G6X..00	Cerebrl infarctn due/unspcf occlusn or sten/cerebrl artrs
5051	G61..00	Intracerebral haemorrhage	**34135**	14A7.00	H/O: CVA/stroke
5254	G340100	Double coronary vessel disease	**34328**	G311300	Refractory angina
5268	G650.11	Insufficiency - basilar artery	**34633**	G34y.00	Other specified chronic ischaemic heart disease
5387	G301.00	Other specified anterior myocardial infarction	**34803**	G30y.00	Other acute myocardial infarction
5413	G340.00	Coronary atherosclerosis	**35674**	14A3.00	H/O: myocardial infarct <60
5871	14A7.12	H/O: stroke	**35713**	G34yz00	Other specified chronic ischaemic heart disease NOS
6116	G66..13	CVA - Cerebrovascular accident unspecified	**36423**	G36..00	Certain current complication follow acute myocardial infarct
6253	G66..12	Stroke unspecified	**36523**	G311.00	Preinfarction syndrome
6960	G61..11	CVA - cerebrovascular accid due to intracerebral haemorrhage	**36609**	G342.00	Atherosclerotic cardiovascular disease
7320	G343.00	Ischaemic cardiomyopathy	**37657**	G362.00	Ventric septal defect/curr comp fol acut myocardal infarctn
7347	G311100	Unstable angina	**38609**	G351.00	Subsequent myocardial infarction of inferior wall

7696	G33z200	Syncope anginosa	**39344**	G676000	Cereb infarct due cerebral venous thrombosis, nonpyogenic
7780	G667.00	Left sided CVA	**39449**	G312.00	Coronary thrombosis not resulting in myocardial infarction
7912	G614.00	Pontine haemorrhage	**39546**	Gyu3000	[X]Other forms of angina pectoris
8443	G663.00	Brain stem stroke syndrome	**39655**	G311.12	Impending infarction
8935	G302.00	Acute inferolateral infarction	**39693**	G31y200	Subendocardial ischaemia
9276	G31y000	Acute coronary insufficiency	**40338**	G611.00	Internal capsule haemorrhage
9413	G31y.00	Other acute and subacute ischaemic heart disease	**40399**	14A4.00	H/O: myocardial infarct >60
9507	G307000	Acute non-Q wave infarction	**40429**	G301000	Acute anteroapical infarction
9555	G33z500	Post infarct angina	**40758**	G6W..00	Cereb infarct due unsp occlus/stenos precerebr arteries
10562	G307100	Acute non-ST segment elevation myocardial infarction	**41221**	G30y200	Acute septal infarction
10794	G656.00	Vertebrobasilar insufficiency	**41835**	G384.00	Postoperative subendocardial myocardial infarction
11983	G311500	Acute coronary syndrome	**44765**	G653.00	Carotid artery syndrome hemispheric
12139	G300.00	Acute anterolateral infarction	**45476**	14AL.00	H/O: Treatment for ischaemic heart disease
12229	G30X000	Acute ST segment elevation myocardial infarction	**45809**	G350.00	Subsequent myocardial infarction of anterior wall
12804	G33z700	Stable angina	**46017**	G30yz00	Other acute myocardial infarction NOS
12833	G668.00	Right sided CVA	**46112**	G380.00	Postoperative transmural myocardial infarction anterior wall
13564	G613.00	Cerebellar haemorrhage	**46166**	G35X.00	Subsequent myocardial infarction of unspecified site
13566	G30..11	Attack - heart	**46276**	G381.00	Postoperative transmural myocardial infarction inferior wall
13571	G30..16	Thrombosis - coronary	**46316**	G612.00	Basal nucleus haemorrhage
14658	G30z.00	Acute myocardial infarction NOS	**47637**	Gyu3300	[X]Other forms of chronic ischaemic heart disease
14897	G301z00	Anterior myocardial infarction NOS	**50372**	14AH.00	H/O: Myocardial infarction in last year
14898	G305.00	Lateral myocardial infarction NOS	**50594**	G654.00	Multiple and bilateral precerebral artery syndromes
15661	G310.11	Dressler's syndrome	**51767**	G666.00	Pure sensory lacunar syndrome
15754	G34z.00	Other chronic ischaemic heart disease NOS	**52517**	Gyu3.00	[X]Ischaemic heart diseases
15788	G65zz00	Transient cerebral ischaemia NOS	**53745**	Gyu6400	[X]Other cerebral infarction
16408	G32..11	Healed myocardial infarction	**53810**	Gyu6200	[X]Other intracerebral haemorrhage
16507	G65z100	Intermittent cerebral ischaemia	**54251**	G311z00	Preinfarction syndrome NOS
17307	G311200	Angina at rest	**54535**	G33z100	Stenocardia
17322	G664.00	Cerebellar stroke syndrome	**55137**	G311011	MI - myocardial infarction aborted
17464	G32..12	Personal history of myocardial infarction	**55247**	G65z000	Impending cerebral ischaemia
17689	G30..17	Silent myocardial infarction	**57315**	G618.00	Intracerebral haemorrhage, multiple localized
17872	G301100	Acute anteroseptal infarction	**59189**	G363.00	Ruptur cardiac wall w'out haemopericard/cur comp fol ac MI
18118	G311400	Worsening angina	**59940**	G364.00	Ruptur chordae tendinae/curr comp fol acute myocard infarct
18125	G330000	Nocturnal angina	**61072**	G311000	Myocardial infarction aborted
18604	G61..12	Stroke due to intracerebral haemorrhage	**62342**	G615.00	Bulbar haemorrhage
18689	G660.00	Middle cerebral artery syndrome	**62626**	G30y100	Acute papillary muscle infarction
18842	G35..00	Subsequent myocardial infarction	**63467**	G306.00	True posterior myocardial infarction
18889	G34z000	Asymptomatic coronary heart disease	**63746**	Fyu5500	[X]Other transnt cerebral ischaemic attacks+related syndroms
19201	G61X100	Right sided intracerebral haemorrhage, unspecified	**66388**	G33z000	Status anginosus
19260	G662.00	Posterior cerebral artery syndrome	**68357**	G31y100	Microinfarction of heart
19280	G661.00	Anterior cerebral artery syndrome	**68401**	Gyu3200	[X]Other forms of acute ischaemic heart disease
19348	ZV12511	[V]Personal history of stroke	**68748**	G38z.00	Postoperative myocardial infarction, unspecified
19354	G65y.00	Other transient cerebral ischaemia	**69474**	G365.00	Rupture papillary muscle/curr comp fol acute myocard infarct

19655	G311.14	Angina at rest	**72562**	G353.00	Subsequent myocardial infarction of other sites
20095	G330.00	Angina decubitus	**90572**	Gyu6500	[X]Occlusion and stenosis of other precerebral arteries
20416	G3...12	Atherosclerotic heart disease	**91627**	Gyu6300	[X]Cerebrl infarctn due/unspcf occlusn or sten/cerebrl artrs
21118	G651000	Vertebro-basilar artery syndrome	**92036**	Gyu6600	[X]Occlusion and stenosis of other cerebral arteries
21844	G31y300	Transient myocardial ischaemia	**94482**	Gyu6G00	[X]Cereb infarct due unsp occlus/stenos precerebr arteries
22383	G3y..00	Other specified ischaemic heart disease	**96630**	Gyu6F00	[X]Intracerebral haemorrhage in hemisphere, unspecified
23078	G34y100	Chronic myocardial ischaemia	**96838**	Gyu3400	[X]Acute transmural myocardial infarction of unspecif site
23465	G652.00	Subclavian steal syndrome	**99991**	Gyu3600	[X]Subsequent myocardial infarction of unspecified site
23579	G310.00	Postmyocardial infarction syndrome	**100139**	14AT.00	History of myocardial infarction
23671	G63y000	Cerebral infarct due to thrombosis of precerebral arteries	**101251**	ZV12D00	[V]Personal history of transient ischaemic attack
23708	G361.00	Atrial septal defect/curr comp folow acut myocardal infarct	**105216**	14AW.00	H/O acute coronary syndrome
23892	G304.00	Posterior myocardial infarction NOS	**105479**	G39..00	Coronary microvascular disease
23942	G650.00	Basilar artery syndrome	**105738**	G657.00	Carotid territory transient ischaemic attack
24126	G360.00	Haemopericardium/current comp folow acut myocard infarct	**106812**	G383.00	Postoperative transmural myocardial infarction unspec site
24446	G63y100	Cerebral infarction due to embolism of precerebral arteries	**107440**	G619.00	Lobar cerebral haemorrhage
24540	G34y000	Chronic coronary insufficiency	**109035**	Gyu3500	[X]Subsequent myocardial infarction of other sites

Codelist for MI (occurrence of event)

medcode	readcode	readterm	medcode	readcode	readterm
241	G30..00	Acute myocardial infarction	**30421**	G30..13	Cardiac rupture following myocardial infarction (MI)
1204	G30..14	Heart attack	**32272**	G38..00	Postoperative myocardial infarction
1677	G30..15	MI - acute myocardial infarction	**32854**	G30B.00	Acute posterolateral myocardial infarction
1678	G308.00	Inferior myocardial infarction NOS	**34803**	G30y.00	Other acute myocardial infarction
2491	G30..12	Coronary thrombosis	**35119**	G501.00	Post infarction pericarditis
3704	G307.00	Acute subendocardial infarction	**36423**	G36..00	Certain current complication follow acute myocardial infarct
5387	G301.00	Other specified anterior myocardial infarction	**37657**	G362.00	Ventric septal defect/curr comp fol acut myocardal infarctn
8935	G302.00	Acute inferolateral infarction	**38609**	G351.00	Subsequent myocardial infarction of inferior wall
9507	G307000	Acute non-Q wave infarction	**40429**	G301000	Acute anteroapical infarction
10562	G307100	Acute non-ST segment elevation myocardial infarction	**41221**	G30y200	Acute septal infarction
12139	G300.00	Acute anterolateral infarction	**41835**	G384.00	Postoperative subendocardial myocardial infarction
12229	G30X000	Acute ST segment elevation myocardial infarction	**45809**	G350.00	Subsequent myocardial infarction of anterior wall
13566	G30..11	Attack - heart	**46017**	G30yz00	Other acute myocardial infarction NOS
13571	G30..16	Thrombosis - coronary	**46112**	G380.00	Postoperative transmural myocardial infarction anterior wall
14658	G30z.00	Acute myocardial infarction NOS	**46166**	G35X.00	Subsequent myocardial infarction of unspecified site
14897	G301z00	Anterior myocardial infarction NOS	**46276**	G381.00	Postoperative transmural myocardial infarction inferior wall
14898	G305.00	Lateral myocardial infarction NOS	**59189**	G363.00	Ruptur cardiac wall w'out haemopericard/cur comp fol ac MI
15661	G310.11	Dressler's syndrome	**59940**	G364.00	Ruptur chordae tendinae/curr comp fol acute myocard infarct

17133	G30A.00	Mural thrombosis	**61670**	889A.00	Diab mellit insulin-glucose infus acute myocardial infarct
17689	G30..17	Silent myocardial infarction	**62626**	G30y100	Acute papillary muscle infarction
17872	G301100	Acute anteroseptal infarction	**63467**	G306.00	True posterior myocardial infarction
18842	G35..00	Subsequent myocardial infarction	**68357**	G31y100	Microinfarction of heart
23708	G361.00	Atrial septal defect/curr comp folow acut myocardal infarct	**68748**	G38z.00	Postoperative myocardial infarction, unspecified
23892	G304.00	Posterior myocardial infarction NOS	**69474**	G365.00	Rupture papillary muscle/curr comp fol acute myocard infarct
24126	G360.00	Haemopericardium/current comp folow acut myocard infarct	**72562**	G353.00	Subsequent myocardial infarction of other sites
28736	G30y000	Acute atrial infarction	**96838**	Gyu3400	[X]Acute transmural myocardial infarction of unspecif site
29553	G366.00	Thrombosis atrium,auric append&vent/curr comp foll acute MI	**99991**	Gyu3600	[X]Subsequent myocardial infarction of unspecified site
29643	G303.00	Acute inferoposterior infarction	**106812**	G383.00	Postoperative transmural myocardial infarction unspec site
29758	G30X.00	Acute transmural myocardial infarction of unspecif site	**107848**	Fyu4000	[X]Other specified acute disseminated demyelination
30330	G309.00	Acute Q-wave infarct	**108103**	K043100	Acute renal failure induced by aminoglycoside
109035	Gyu3500	[X]Subsequent myocardial infarction of other sites	**108530**	Fyu4200	[X]Acute disseminated demyelination, unspecified

Stroke (occurrence of event)

medcode	readcode	readterm	medcode	readcode	readterm
504	G65..00	Transient cerebral ischaemia	**23671**	G63y000	Cerebral infarct due to thrombosis of precerebral arteries
1298	G66..11	CVA unspecified	**23942**	G650.00	Basilar artery syndrome
1433	G65..12	Transient ischaemic attack	**24446**	G63y100	Cerebral infarction due to embolism of precerebral arteries
1469	G66..00	Stroke and cerebrovascular accident unspecified	**28314**	G61X000	Left sided intracerebral haemorrhage, unspecified
1895	G65z.00	Transient cerebral ischaemia NOS	**30045**	G616.00	External capsule haemorrhage
2417	G65..13	Vertebro-basilar insufficiency	**31060**	G61X.00	Intracerebral haemorrhage in hemisphere, unspecified
3132	G65..11	Drop attack	**31595**	G610.00	Cortical haemorrhage
3535	G61z.00	Intracerebral haemorrhage NOS	**33377**	G651.00	Vertebral artery syndrome
5051	G61..00	Intracerebral haemorrhage	**33499**	G665.00	Pure motor lacunar syndrome
5268	G650.11	Insufficiency - basilar artery	**33543**	G6X..00	Cerebrl infarctn due/unspcf occlusn or sten/cerebrl artrs
6116	G66..13	CVA - Cerebrovascular accident unspecified	**39344**	G676000	Cereb infarct due cerebral venous thrombosis, nonpyogenic
6253	G66..12	Stroke unspecified	**40338**	G611.00	Internal capsule haemorrhage
6960	G61..11	CVA - cerebrovascular accid due to intracerebral haemorrhage	**40758**	G6W..00	Cereb infarct due unsp occlus/stenos precerebr arteries
7780	G667.00	Left sided CVA	**44765**	G653.00	Carotid artery syndrome hemispheric
7912	G614.00	Pontine haemorrhage	**46316**	G612.00	Basal nucleus haemorrhage
8443	G663.00	Brain stem stroke syndrome	**50594**	G654.00	Multiple and bilateral precerebral artery syndromes
10794	G656.00	Vertebrobasilar insufficiency	**51767**	G666.00	Pure sensory lacunar syndrome
12833	G668.00	Right sided CVA	**53745**	Gyu6400	[X]Other cerebral infarction
13564	G613.00	Cerebellar haemorrhage	**53810**	Gyu6200	[X]Other intracerebral haemorrhage
15788	G65zz00	Transient cerebral ischaemia NOS	**55247**	G65z000	Impending cerebral ischaemia
16507	G65z100	Intermittent cerebral ischaemia	**57315**	G618.00	Intracerebral haemorrhage, multiple localized
17322	G664.00	Cerebellar stroke syndrome	**62342**	G615.00	Bulbar haemorrhage
18604	G61..12	Stroke due to intracerebral haemorrhage	**63746**	Fyu5500	[X]Other transnt cerebral ischaemic attacks+related syndroms
18689	G660.00	Middle cerebral artery syndrome	**90572**	Gyu6500	[X]Occlusion and stenosis of other precerebral arteries
19201	G61X100	Right sided intracerebral haemorrhage, unspecified	**91627**	Gyu6300	[X]Cerebrl infarctn due/unspcf occlusn or sten/cerebrl artrs
19260	G662.00	Posterior cerebral artery syndrome	**92036**	Gyu6600	[X]Occlusion and stenosis of other cerebral arteries

19280	G661.00	Anterior cerebral artery syndrome	**94482**	Gyu6G00	[X]Cereb infarct due unsp occlus/stenos precerebr arteries
19354	G65y.00	Other transient cerebral ischaemia	**96630**	Gyu6F00	[X]Intracerebral haemorrhage in hemisphere, unspecified
21118	G651000	Vertebro-basilar artery syndrome	**105738**	G657.00	Carotid territory transient ischaemic attack
23465	G652.00	Subclavian steal syndrome	**107440**	G619.00	Lobar cerebral haemorrhage

BNF codes for concomitant medications

Statins: 02120400.

Anti-hypertensive medications: any thiazide diuretic (02020100) (plus any combinations including this), any calcium channel blocker (02060200), Angiotensin-converting enzyme (ACE) inhibitors (02050501), Angiotensin-II receptor antagonists (02050502) and Beta Blockers (02040000).

Non-steroidal anti-inflammatory medications (NSAIDs):100101

Aspirin: Any code starting 0209* plus a partial string match of *Asp

Smoking

Smoking status is a specific entry field in the CPRD, with data stored in the "additional" file under a specific type (entity type 4). The raw entry in CPRD classifies the smoking status into not entered, Smoker, Non-smoker and ex-smoker. All available entries were extracted for a patient. For multiple entries on the same day, if any "not entered" appeared alongside another code, then the not entered code was deleted. Any other discrepancies on the same day were changed to missing ("not entered"). Differences in coding were then examined longitudinally prior to the date of diabetes diagnosis. The closest entry prior to the date of diabetes diagnosis was examined. If this was current or ex-smoker, it was kept as such. If it was non-smoker, then it was only kept as non-smoker if all entries prior to this entry (if any) were also non-smoker or unknown. If any previous entry suggested the patient was a current or ex-smoker, they were changed to be an ex-smoker. This value was then taken as the fixed smoking status of the patient at the time of diabetes diagnosis and was not time updated after this point.

Alcohol

Alcohol consumption was determined via use of read codes (listed below) from both the clinical and referral files. Information on reported daily and weekly alcohol consumption was obtained from the additional information file where it is entered under entity type 5.

The pre-existing code list had been manually sorted into 7 simple categories of “non-drinker”, “ex drinker”, “rare drinker”, “current drinker”, “excessive drinker”, and “drinker but amount not specified” and “missing”.

If available, the more detailed information on daily and weekly units were also used. The threshold of units per day or week to classify rare, moderate and excessive drinking was applied differently for men and women based on standard UK guidelines. Based on weekly information, rare was <1 unit per week for males and females, current was 1-21 (inclusive) units per week for males and 1-14 (inclusive) units per week for females. Excessive was more than 21 units per week for males and >14 units per week for females.

Daily information was classified such that <2 units per day was rare, 3-6 units per day as moderate, and >6 units per day as excessive, for both males and females.

All three classifications were then considered together to calculate the most appropriate category for each patient on each date. Discrepancies between clinical, weekly and daily info were dealt with a simple rule. As with smoking, a non-missing entry took priority over a missing entry. If there were two non “unknown” entries on the same day, the value taken was the worst case.

The closest record to the time of diabetes onset (but before) was then identified and any prior entries that were the same alcohol category were deleted for simplicity. A similar process to the smoking variable was then employed. Specifically, if the closest code to diagnosis was non-drinker, but ex-drinker appeared previously, they were considered an ex-drinker. If either of current or rare drinker appeared previously, they were considered a rare drinker.

medcode	readcode	readterm	Category
27	136..00	Alcohol consumption	Non specified drinker
385	1362.11	Drinks rarely	Rare drinker
669	E250000	Nondependent alcohol abuse, unspecified	Excessive Drinker
749	1362.12	Drinks occasionally	Rare drinker
956	136J.00	Social drinker	Current Drinker
967	1367	Stopped drinking alcohol	Ex drinker
1399	E23..12	Alcohol problem drinking	Excessive Drinker
2689	136G.00	Beer drinker	Current Drinker
3782	E250.14	Intoxication - alcohol	Current Drinker
4447	1361.12	Non-drinker alcohol	Non drinker
7545	ZV4KC00	[V] Alcohol use	Non specified drinker
7746	E250.00	Nondependent alcohol abuse	Excessive Drinker
8999	136P.00	Heavy drinker	Excessive Drinker
9169	R103.00	[D]Alcohol blood level excessive	Current Drinker
10161	2577.11	O/E - alcoholic breath	Current Drinker
12271	E250.11	Drunkenness NOS	Current Drinker
12949	1361	Teetotaller	Non drinker
12968	136H.00	Drinks beer and spirits	Current Drinker
12969	136I.00	Drinks wine	Current Drinker
12970	1361.11	Non drinker alcohol	Non drinker
12971	136F.00	Spirit drinker	Current Drinker
12974	E250200	Nondependent alcohol abuse, episodic	Excessive Drinker
12978	1368	Alcohol consumption unknown	Non specified drinker

12979	136M.00	Current non drinker	Non drinker
12980	136N.00	Light drinker	Current Drinker
12981	136Z.00	Alcohol consumption NOS	Non specified drinker
12982	136K.00	Alcohol intake above recommended sensible limits	Excessive Drinker
12983	136E.00	Ex-very heavy drinker-(>9u/d)	Ex drinker
12984	136Q.00	Very heavy drinker	Excessive Drinker
12985	136O.00	Moderate drinker	Current Drinker
16587	ZV11311	[V]Problems related to lifestyle alcohol use	Excessive Drinker
17777	E250.13	Inebriety NOS	Current Drinker
19401	136R.00	Binge drinker	Excessive Drinker
19493	136D.00	Ex-heavy drinker - (7-9u/day)	Ex drinker
19494	136S.00	Hazardous alcohol use	Excessive Drinker
19495	136C.00	Ex-moderate drinker - (3-6u/d)	Ex drinker
21829	Z786200	Drinking practice	Non specified drinker
22933	136A.00	Ex-trivial drinker (<1u/day)	Ex drinker
23610	E250100	Nondependent alcohol abuse, continuous	Excessive Drinker
23978	U81..00	[X]Evid of alcohol involv determind by level of intoxication	Current Drinker
24735	2577	O/E - breath - alcohol smell	Current Drinker
26471	136B.00	Ex-light drinker - (1-2u/day)	Ex drinker
26472	136L.00	Alcohol intake within recommended sensible limits	Current Drinker
27518	E250.12	Hangover (alcohol)	Current Drinker
28150	E250z00	Nondependent alcohol abuse NOS	Excessive Drinker
30695	136T.00	Harmful alcohol use	Excessive Drinker
31569	E250300	Nondependent alcohol abuse in remission	Excessive Drinker
44783	1D19.00	Pain in lymph nodes after alcohol consumption	Current Drinker
84218	13ZY.00	Disqualified from driving due to excess alcohol	Excessive Drinker
93415	136V.00	Alcohol units per week	Non specified drinker
94670	136W.00	Alcohol misuse	Excessive Drinker
95944	9k19.00	Alcohol assesment declined - enhanced services admin	Non specified drinker
96259	9k19.11	Alcohol assessment declined	Non specified drinker
97126	136X.00	Alcohol units consumed on heaviest drinking day	Non specified drinker

Chronic Kidney disease

medcode	readcode	readterm
512	K05..00	Chronic renal failure
6712	K050.00	End stage renal failure
8330	K0D..00	End-stage renal disease
12479	1Z13.00	Chronic kidney disease stage 4
12566	1Z12.00	Chronic kidney disease stage 3
12585	1Z14.00	Chronic kidney disease stage 5
19473	66i..00	Chronic kidney disease monitoring
53852	K05..12	End stage renal failure
94793	1Z1B.00	Chronic kidney disease stage 3 with proteinuria
94965	1Z15.00	Chronic kidney disease stage 3A
95122	1Z1H.00	Chronic kidney disease stage 4 with proteinuria
95123	1Z1C.00	Chronic kidney disease stage 3 without proteinuria
95175	1Z1E.00	Chronic kidney disease stage 3A without proteinuria
95177	1Z1G.00	Chronic kidney disease stage 3B without proteinuria
95178	1Z1F.00	Chronic kidney disease stage 3B with proteinuria
95179	1Z16.00	Chronic kidney disease stage 3B
95405	1Z1L.00	Chronic kidney disease stage 5 without proteinuria
95406	1Z1J.00	Chronic kidney disease stage 4 without proteinuria
95408	1Z1D.00	Chronic kidney disease stage 3A with proteinuria
95508	1Z1K.00	Chronic kidney disease stage 5 with proteinuria
104619	K053.00	Chronic kidney disease stage 3
104963	K054.00	Chronic kidney disease stage 4
104981	K05..13	Chronic kidney disease
105151	K055.00	Chronic kidney disease stage 5

Systolic Blood Pressure (SBP)

SBP was extracted from the additional information file by searching entity type 1 and taking the data from the systolic data entry field. SBP of < 80 or > 200 were set to missing as they were considered implausible values. Specifically, a SBP of < 80 may be suggestive that DBP had been entered into the wrong field. A SBP of greater than 180 indicates the need for emergency care [215] and is unlikely to be measured in a routine clinical visit, however a cut off of 200 was used to allow for this possibility within a diabetic population. This resulted in excluding less than 1% of all measurements. Duplicate entries on the same day were averaged if the absolute difference between the highest and lowest was less than 10, otherwise they were set to missing. The set of longitudinal measures for each patients was saved.

BMI

For the final cohort of patients, height and weight were extracted by searching the additional information file for entity types 13 and 14. These were then merged with the clinical information file to get dates of measurements. Height was cleaned by first searching for values >100, suggesting height had been entered in cm. These were converted to m. After this, entries suggesting implausible values (>2.3 m (7.5 feet) and <1.3 m (4 feet)) were set to missing. Multiple height entries on the same day were replaced with the mean if the difference was less than 5 cm, or both created missing. Weight was then cleaned in a similar way, with implausible values removed (<4 stone (25kg) or > 40 stone (255kg)), and multiple entries on the same day replaced with means, or removed if the differences between the two results differed by 2kg or more.

BMI was then calculated using the cleaned height and weight measures, and the set of longitudinal measured for each patient saved.

HbA1c (%)

HbA1c records, consisting of the value recorded, unit of measurement and date of record, were extracted from the test file by searching entity type 275. Results were only kept if they were entered as %, mmol/L, iu/L and mmol/mol. (this accounted for 80% of all entries, with a further 19% being labelled as “unit not entered”). All results were then converted to %. After conversion, readings of <2% and >20 % were considered implausible. Duplicates on the same day were then identified. Any result originally in % format was given priority. Multiple readings in % were then compared and removed if the difference was > 2%. Otherwise a mean was used.

APPENDIX 13 CKD STAGES AND TREATMENT INITIATION

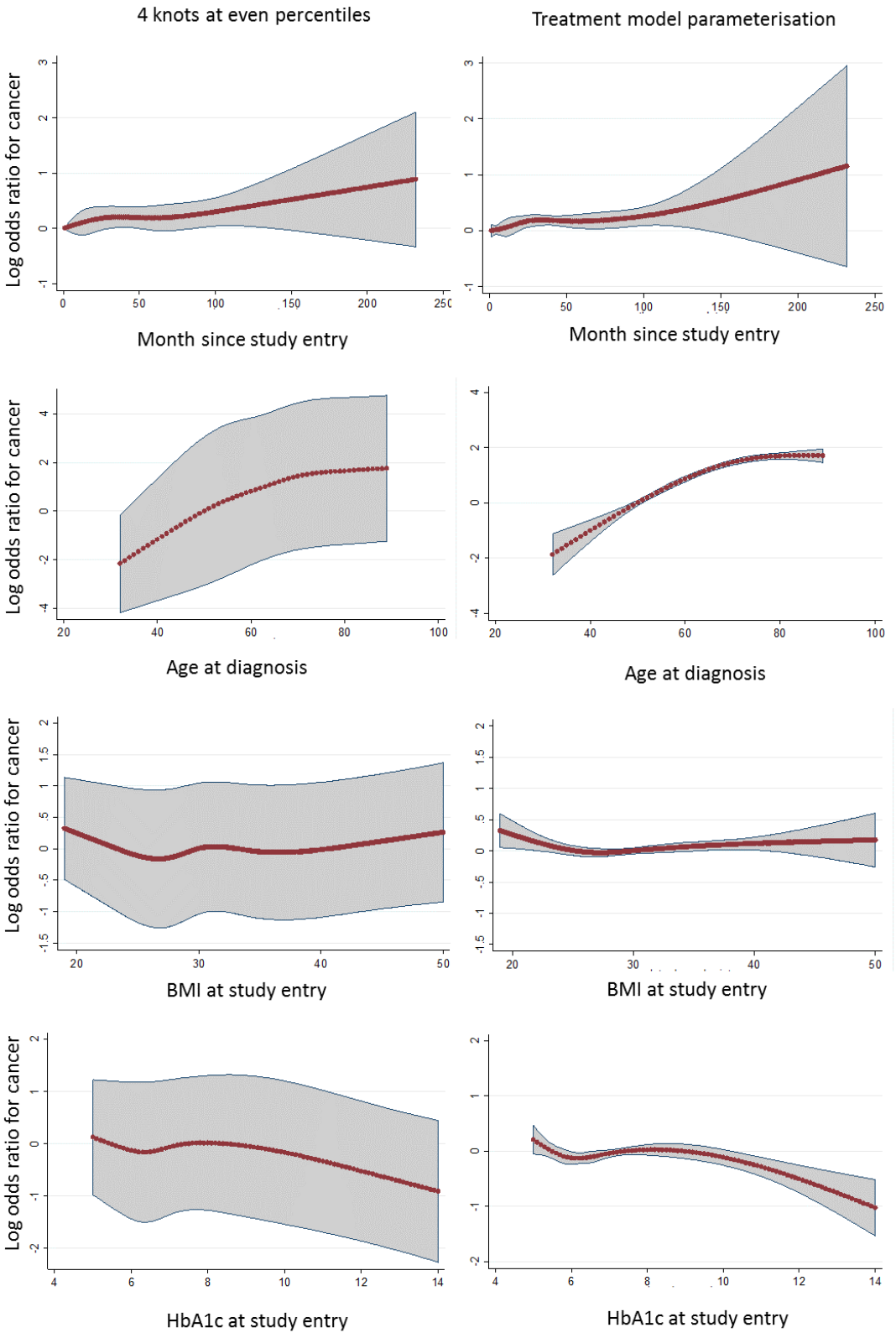
The following tables show the number (top) and percent (bottom) of person months up to and including the interval of treatment initiation, that fall within each stage of CKD.

Treatment initiated	No CKD	No stage indicated	Stage 3	Stage 4	Stage 5	Total
None	1,295,044	11,329	202,060	4,423	582	1,513,438
Metformin	23,201	81	1,626	9	3	24,920
Sulfonylurea	1,963	32	287	25	4	2,311

Treatment initiated	No CKD	No stage indicated	Stage 3	Stage 4	Stage 5	Total
None	85.57%	0.75%	13.35%	0.29%	0.04%	100%
Metformin	93.10%	0.33%	6.52%	0.04%	0.01%	100%
Sulfonylurea	84.94%	1.38%	12.42%	1.08%	0.17%	100%

As shown, Overall, the proportion of patients that are at any time classified as stage 4 or stage 5 is small, however, it is particularly noticeable when looking at proportions, that patients with stage 4 and stage 5 are substantially less likely to initiate metformin over a sulfonylurea.

APPENDIX 14 COMPARISON OF SPLINE PARAMETERISATIONS OF BASELINE COVARIATES FOR RISK OF OUTCOME



APPENDIX 15 TREATMENT AND CENSORING MODEL OUTPUT FOR FOLLOW UP TO CANCER

Treatment models

Table 15.1 Estimated OR, standard error and 95% CI for probability of treatment with metformin for denominator and numerator models for the IPTW. Covariate specification A

	DENOMINATOR MODEL			NUMERATOR MODEL		
	OR	SE	95% CI	OR	SE	95% CI
BASELINE FIXED COVARIATES						
Time (months) since study entry						
Time since study entry (months) spl 1	0.5	0.01	0.48 , 0.53	0.4	0.01	0.38 , 0.41
Time since study entry (months) spl 2	1.61	0.04	1.53 , 1.69	2.36	0.06	2.25 , 2.47
Time since study entry (months) spl 3	0.62	0.02	0.59 , 0.65	0.34	0.01	0.33 , 0.36
Time since study entry (months) spl 4	1.2	0.02	1.17 , 1.24	1.48	0.02	1.44 , 1.52
Time (months) between diagnosis and study entry						
Time between diagnosis and study entry spl 1	0.87	0.02	0.82 , 0.92	0.85	0.03	0.8 , 0.91
Time between diagnosis and study entry spl 2	1.07	0.02	1.03 , 1.12	0.85	0.02	0.81 , 0.89
Age at diagnosis (years)						
Age at diagnosis spl1	0.89	0.01	0.87 , 0.91	0.8	0.01	0.79 , 0.82
Age at diagnosis spl2	0.92	0.01	0.9 , 0.94	0.91	0.01	0.9 , 0.93
Age at diagnosis spl3	1.04	0.01	1.02 , 1.05	1.01	0.01	0.99 , 1.03
Gender (FvM(	1.15	0.02	1.11 , 1.19	1.12	0.02	1.08 , 1.16
Smoking Status						
Non	1 (ref)			1 (ref)		
Current	1.03	0.03	0.98 , 1.08	1.09	0.02	1.04 , 1.13
Ex	1.07	0.02	1.03 , 1.11	1.07	0.02	1.03 , 1.1
Alcohol consumption						
non_drinker	1 (ref)			1 (ref)		
ex-drinker	0.94	0.04	0.86 , 1.02	0.95	0.04	0.88 , 1.02
current drinker unknown	0.80	0.06	0.69 , 0.93	0.87	0.05	0.77 , 0.98
rare drinker <2u/d	0.96	0.03	0.91 , 1.02	0.97	0.03	0.92 , 1.02
moderate drinker 3-6u/d	0.96	0.03	0.91 , 1.02	0.92	0.02	0.88 , 0.97
excessive drinker >6u/d	0.92	0.03	0.86 , 0.99	0.86	0.03	0.8 , 0.92
Year of diabetes onset						
1990-1994	1 (ref)			1 (ref)		
1995-2000	1.39	0.32	0.88 , 2.2	1.16	0.23	0.79 , 1.71
2001-2005	1.46	0.34	0.93 , 2.3	1.09	0.21	0.74 , 1.59
2005 onwards	1.52	0.35	0.97 , 2.4	1.05	0.2	0.72 , 1.54
Use of anti HT in year prior to study entry	0.86	0.03	0.81 , 0.93	0.98	0.02	0.95 , 1.01
Use of statin in year prior to study entry	0.86	0.02	0.82 , 0.9	1.2	0.02	1.16 , 1.24
Use of NSAID in year prior to study entry	1.08	0.03	1.03 , 1.13	1.16	0.02	1.12 , 1.2
HbA1c at study entry (%)						
HbA1c at study entry spl1	0.72	0.02	0.7 , 0.75	2.33	0.03	2.27 , 2.39
HbA1c at study entry spl2	1.25	0.02	1.22 , 1.29	0.67	0.01	0.65 , 0.68
HbA1c at study entry spl3	0.9	0.01	0.88 , 0.92	1.11	0.01	1.09 , 1.13
HbA1c at study entry spl4	1.08	0.01	1.06 , 1.1	1.06	0.01	1.05 , 1.08
BMI at study entry (kg/m^2)						
BMI at study entry spl1	0.77	0.03	0.72 , 0.83	1.16	0.01	1.14 , 1.18
BMI at study entry spl2	0.94	0.01	0.93 , 0.96	0.92	0.01	0.9 , 0.94
BMI at study entry spl3	1.04	0.01	1.02 , 1.06	1.04	0.01	1.02 , 1.06
History of CVD at study entry	0.94	0.06	0.83 , 1.06	1.05	0.02	1.01 , 1.1
History of CKD at study entry	0.81	0.05	0.72 , 0.91	0.72	0.03	0.67 , 0.78
TIME UPDATED COVARIATES						
Use of anti HT in previous year	1.08	0.03	1.03 , 1.13			
Use of statin in previous year	1.65	0.04	1.57 , 1.73			
Use of NSAID in previous year	1.22	0.04	1.14 , 1.31			
History of CVD	1.02	0.06	0.91 , 1.15			
History of CKD	0.9	0.04	0.83 , 0.98			
HbA1c in previous interval (%)						
Previous HbA1c spl1	5.48	0.09	5.31 , 5.66			
Previous HbA1c spl2	0.5	0.01	0.48 , 0.51			
Previous HbA1c spl3	1.48	0.02	1.43 , 1.52			
BMI in previous interval (per kg/m^2 increase)	1.07	0.01	1.06 , 1.08			

Table 15.2 Estimated OR, standard error and 95% CI for probability of treatment with metformin for denominator and numerator models for the IPTW, using covariate specification B.

	DENOMINATOR MODEL			NUMERATOR MODEL		
	OR	SE	95% CI	OR	SE	95% CI
BASELINE FIXEDCOVARIATES						
Time (months) since study entry						
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 , 1.05	1.08	0.004	1.07 , 1.09
Time (months) between diagnosis and study entry						
Time between diagnosis and study entry spl 1	0.87	0.02	0.82 , 0.92	0.85	0.03	0.8 , 0.9
Time between diagnosis and study entry spl 2	1.09	0.02	1.04 , 1.13	0.85	0.02	0.81 , 0.89
Age at diagnosis (years)						
Age at diagnosis spl1	0.89	0.01	0.88 , 0.91	0.8	0.01	0.79 , 0.82
Age at diagnosis spl2	0.92	0.01	0.9 , 0.93	0.91	0.01	0.89 , 0.92
Age at diagnosis spl3	1.04	0.01	1.03 , 1.06	1.01	0.01	1 , 1.03
Gender (FvM)	1.15	0.02	1.11 , 1.19	1.11	0.02	1.08 , 1.15
Smoking Status						
Non	1 (ref)					
Current	1.03	0.03	0.98 , 1.08	1.09	0.03	1.04 , 1.14
Ex	1.08	0.02	1.04 , 1.12	1.08	0.02	1.04 , 1.11
Alcohol consumption						
non_drinker	1 (ref)					
ex-drinker	0.93	0.04	0.85 , 1.02	0.95	0.04	0.88 , 1.02
current drinker unknown	0.79	0.06	0.68 , 0.92	0.86	0.05	0.77 , 0.98
rare drinker <2u/d	0.97	0.03	0.91 , 1.03	0.96	0.03	0.91 , 1.02
moderate drinker 3-6u/d	0.97	0.03	0.91 , 1.02	0.92	0.02	0.88 , 0.97
excessive drinker >6u/d	0.92	0.03	0.86 , 0.99	0.86	0.03	0.8 , 0.92
Year of diabetes onset						
1990-1994	1 (ref)					
1995-2000	1.4	0.32	0.89 , 2.19	1.15	0.25	0.74 , 1.77
2001-2005	1.51	0.34	0.97 , 2.35	1.07	0.23	0.69 , 1.63
2005 onwards	1.58	0.36	1.02 , 2.47	1.04	0.23	0.68 , 1.59
Use of anti HT in year prior to study entry	0.87	0.03	0.81 , 0.93	0.98	0.02	0.95 , 1.02
Use of statin in year prior to study entry	0.87	0.02	0.83 , 0.92	1.21	0.02	1.17 , 1.26
Use of NSAID in year prior to study entry	1.09	0.03	1.04 , 1.14	1.17	0.02	1.13 , 1.21
HbA1c at study entry						
HbA1c at study entry linear	0.33	0.02	0.29 , 0.37	7.69	0.35	7.03 , 8.41
HbA1c at study quadratic	1.05	0.004	1.05 , 1.06	0.91	0.002	0.91 , 0.92
BMI at study entry	0.95	0.01	0.94 , 0.96	1.02	0.002	1.02 , 1.02
History of CVD at study entry	0.96	0.06	0.85 , 1.08	1.05	0.02	1.01 , 1.1
History of CKD at study entry	0.8	0.05	0.72 , 0.9	0.72	0.03	0.67 , 0.78
TIME UPDATED COVARIATES						
Use of anti HT in previous year	1.08	0.03	1.03 , 1.13			
Use of statin in previous year	1.62	0.04	1.54 , 1.71			
Use of NSAID in previous year	1.21	0.04	1.13 , 1.3			
History of CVD	1	0.06	0.89 , 1.13			
History of CKD	0.91	0.04	0.84 , 0.99			
HbA1c in previous interval (per % increase)						
Previous HbA1c linear	593.95	49.36	504.68 , 699.02			
Previous HbA1c quadratic	0.73	0	0.73 , 0.74			
BMI in previous interval (per kg/m^2 increase)	1.07	0.01	1.06 , 1.09			

Censoring models

Table 15.3 Parameter estimates from multinomial logistic regression model for probability of censoring, denominator model for weights (LEFT) and numerator model for stabilisation (RIGHT) Covariate specification A

	DENOMINATOR MODEL									NUMERATOR MODEL								
	Censoring due initiation of other antidiabetic medication			Censoring due to death			Censoring due to transfer out			Censoring due initiation of other antidiabetic medication			Censoring due to death			Censoring due to transfer out		
	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI
BASELINE FIXED																		
Time (months) since study entry																		
Time since study entry (months) spl 1	0.76	0.03	0.72 , 0.8	1.32	0.07	1.16 , 1.51	1.28	0.05	1.17 , 1.42	0.41	0.02	0.39 , 0.43	1.54	0.06	1.35 , 1.73	1.22	0.05	1.12 , 1.35
Time since study entry (months) spl 2	1.27	0.03	1.2 , 1.34	1.01	0.07	0.88 , 1.16	0.8	0.05	0.73 , 0.9	2.41	0.03	2.27 , 2.53	0.99	0.07	0.86 , 1.14	0.84	0.05	0.76 , 0.93
Time since study entry (months) spl 3	0.72	0.03	0.68 , 0.76	0.99	0.09	0.84 , 1.17	1.35	0.06	1.2 , 1.51	0.35	0.03	0.33 , 0.37	1.01	0.08	0.86 , 1.2	1.28	0.06	1.15 , 1.45
Time since study entry (months) spl 4	1.14	0.02	1.11 , 1.17	1.03	0.04	0.96 , 1.11	0.9	0.03	0.86 , 0.96	1.38	0.02	1.34 , 1.42	1.02	0.04	0.95 , 1.09	0.91	0.03	0.87 , 0.97
Time (months) between diagnosis and study entry																		
Time between diagnosis and study entry spl 1	0.85	0.03	0.8 , 0.9	1.27	0.04	1.19 , 1.36	1.08	0.03	1.02 , 1.16	0.81	0.04	0.75 , 0.88	1.26	0.04	1.17 , 1.35	1.08	0.04	1.02 , 1.16
Time between diagnosis and study entry spl 2	1.04	0.02	1 , 1.08	0.92	0.03	0.87 , 0.99	1.03	0.03	0.98 , 1.08	0.84	0.03	0.79 , 0.89	0.95	0.03	0.89 , 1.01	1.02	0.03	0.97 , 1.08
Age at diagnosis (years)																		
Age at diagnosis spl1	0.93	0.01	0.91 , 0.95	2.48	0.04	2.29 , 2.69	0.9	0.02	0.86 , 0.92	0.83	0.01	0.81 , 0.84	2.56	0.04	2.36 , 2.77	0.89	0.02	0.85 , 0.91
Age at diagnosis spl2	0.97	0.01	0.95 , 0.99	1.08	0.04	1 , 1.16	1.19	0.01	1.15 , 1.22	1.01	0.01	0.99 , 1.02	1.15	0.04	1.07 , 1.23	1.21	0.01	1.17 , 1.23
Age at diagnosis spl3	1.02	0.01	1.01 , 1.04	1.01	0.02	0.97 , 1.05	0.91	0.01	0.89 , 0.94	0.98	0.01	0.96 , 1	1.02	0.02	0.98 , 1.06	0.91	0.01	0.89 , 0.93
Gender	1.08	0.02	1.04 , 1.12	0.7	0.04	0.65 , 0.76	0.95	0.03	0.89 , 1.01	0.96	0.02	0.93 , 1	0.73	0.04	0.68 , 0.79	0.94	0.03	0.89 , 1.01
Smoking Status																		
Non																		
Current	1.04	0.02	0.99 , 1.08	2.32	0.05	2.08 , 2.59	1.06	0.04	0.97 , 1.15	1.09	0.02	1.05 , 1.14	2.41	0.05	2.18 , 2.66	1.06	0.04	0.98 , 1.16

Ex	1.06	0.02	1.02 , 1.11	1.39	0.04	1.27 , 1.52	0.99	0.03	0.93 , 1.06	1.05	0.02	1.01 , 1.08	1.38	0.04	1.26 , 1.51	0.99	0.03	0.92 , 1.06
Alcohol consumption																		
non_drinker																		
ex-drinker	1.06	0.04	0.98 , 1.16	1.32	0.08	1.12 , 1.55	0.94	0.07	0.81 , 1.08	1.02	0.04	0.94 , 1.12	1.35	0.08	1.14 , 1.58	0.93	0.07	0.81 , 1.07
current drinker unknown	1.16	0.06	1.02 , 1.31	1.55	0.12	1.23 , 1.95	1.42	0.1	1.17 , 1.72	1.19	0.06	1.04 , 1.34	1.62	0.11	1.3 , 2.01	1.45	0.1	1.2 , 1.75
rare drinker <2u/d	0.96	0.03	0.9 , 1.02	1.02	0.07	0.9 , 1.16	0.87	0.05	0.78 , 0.96	0.96	0.03	0.9 , 1.01	1.01	0.07	0.89 , 1.15	0.86	0.05	0.78 , 0.95
moderate drinker 3-6u/d	0.98	0.03	0.92 , 1.03	0.94	0.06	0.84 , 1.06	0.86	0.05	0.79 , 0.94	0.96	0.03	0.91 , 1.01	0.93	0.06	0.83 , 1.05	0.85	0.05	0.78 , 0.94
excessive drinker >6u/d	0.94	0.04	0.88 , 1.02	1.49	0.08	1.27 , 1.75	0.89	0.07	0.78 , 1.01	0.89	0.04	0.83 , 0.96	1.49	0.08	1.27 , 1.75	0.87	0.07	0.76 , 0.99
Year of diabetes onset																		
1990-1994																		
1995-2000	0.56	0.17	0.4 , 0.79	1.3	0.23	0.84 , 2.03	0.81	0.24	0.51 , 1.3	0.48	0.16	0.35 , 0.65	1.25	0.23	0.79 , 1.95	0.79	0.24	0.5 , 1.28
2001-2005	0.36	0.17	0.25 , 0.5	1.19	0.23	0.76 , 1.86	0.83	0.24	0.52 , 1.31	0.28	0.16	0.2 , 0.38	1.07	0.23	0.68 , 1.7	0.79	0.24	0.49 , 1.27
2005 onwards	0.28	0.17	0.2 , 0.39	1.12	0.23	0.71 , 1.77	0.86	0.24	0.54 , 1.38	0.2	0.16	0.15 , 0.28	1.02	0.24	0.64 , 1.62	0.84	0.24	0.52 , 1.35
Use of anti HT in year prior to study entry	0.93	0.03	0.89 , 0.98	1.51	0.07	1.32 , 1.73	1.25	0.06	1.12 , 1.39	0.97	0.02	0.93 , 1.01	1.23	0.05	1.13 , 1.36	0.91	0.03	0.86 , 0.98
Use of statin in year prior to study entry	0.92	0.02	0.89 , 0.96	0.96	0.05	0.88 , 1.06	1.13	0.04	1.04 , 1.22	0.94	0.02	0.91 , 0.98	0.84	0.04	0.78 , 0.91	1.01	0.03	0.94 , 1.07
Use of NSAID in year prior to study entry	1.07	0.02	1.02 , 1.12	0.89	0.05	0.8 , 0.98	1.04	0.04	0.96 , 1.13	1.08	0.02	1.04 , 1.13	0.95	0.05	0.86 , 1.04	0.98	0.04	0.91 , 1.06
HbA1c at study entry																		
HbA1c at study entry spl1	1.05	0.01	1.02 , 1.07	1.13	0.04	1.05 , 1.21	1.03	0.02	0.98 , 1.08	1.6	0.01	1.57 , 1.63	1.13	0.03	1.05 , 1.2	1.07	0.02	1.03 , 1.13
HbA1c at study entry spl2	1	0.01	0.98 , 1.02	0.95	0.03	0.9 , 1.01	0.94	0.02	0.9 , 0.98	0.86	0.01	0.84 , 0.88	0.95	0.03	0.9 , 1	0.93	0.02	0.9 , 0.97
HbA1c at study entry spl3	0.96	0.01	0.94 , 0.99	0.98	0.02	0.94 , 1.02	1.02	0.02	0.98 , 1.05	1	0.01	0.98 , 1.02	1.06	0.02	1.02 , 1.11	1.03	0.02	1 , 1.06
HbA1c at study entry spl4	1.04	0.01	1.02 , 1.06	0.98	0.02	0.95 , 1.01	1.02	0.01	0.99 , 1.05	1.04	0.01	1.02 , 1.06	0.98	0.02	0.95 , 1.01	1.02	0.01	0.99 , 1.05
BMI at study entry																		
BMI at study entry spl1	0.74	0.03	0.7 , 0.79	2.03	0.05	1.84 , 2.27	1.07	0.05	0.98 , 1.19	0.91	0.01	0.9 , 0.93	1.02	0.02	0.97 , 1.06	0.93	0.02	0.9 , 0.97
BMI at study entry spl2	1.14	0.01	1.13 , 1.16	1.22	0.02	1.17 , 1.27	1.01	0.02	0.98 , 1.04	1.12	0.01	1.09 , 1.13	1.22	0.02	1.17 , 1.27	1.01	0.02	0.99 , 1.04

BMI at study entry spl3	0.94	0.01	0.93 , 0.96	0.94	0.02	0.9 , 0.97	0.96	0.01	0.93 , 0.99	0.93	0.01	0.91 , 0.95	0.94	0.02	0.91 , 0.97	0.96	0.01	0.93 , 0.99
History of CVD at study entry	0.99	0.05	0.9 , 1.09	0.77	0.07	0.66 , 0.89	0.83	0.09	0.69 , 0.98	1.17	0.03	1.12 , 1.25	1.67	0.04	1.54 , 1.82	1.07	0.05	0.98 , 1.17
History of CKD at study entry	1.13	0.06	1.01 , 1.27	1.07	0.08	0.92 , 1.26	0.88	0.1	0.73 , 1.06	1.35	0.05	1.22 , 1.48	1.34	0.07	1.17 , 1.52	0.87	0.08	0.75 , 1.02
TIME UPDATED																		
Medication in previous interval	2.44	0.02	2.34 , 2.56	0.9	0.05	0.81 , 0.99	0.87	0.04	0.81 , 0.94	4.53	0.02	4.31 , 4.71	0.87	0.05	0.79 , 0.96	0.85	0.04	0.79 , 0.91
Use of anti HT in previous year	1.08	0.02	1.03 , 1.13	1.15	0.05	1.03 , 1.27	0.87	0.04	0.79 , 0.95									
Use of statin in previous year	1.16	0.02	1.11 , 1.21	0.75	0.05	0.68 , 0.83	0.8	0.04	0.74 , 0.87									
Use of NSAID in previous year	1.17	0.03	1.12 , 1.23	0.7	0.07	0.61 , 0.81	0.66	0.06	0.59 , 0.75									
History of CVD	1.15	0.05	1.04 , 1.26	2.39	0.07	2.1 , 2.75	1.32	0.08	1.13 , 1.55									
History of CKD	1.35	0.04	1.26 , 1.45	1.38	0.05	1.25 , 1.54	1.02	0.06	0.9 , 1.15									
HbA1c in previous interval																		
Previous HbA1c spl1	3.46	0.01	3.35 , 3.56	1.06	0.03	1 , 1.13	1.09	0.02	1.05 , 1.14									
Previous HbA1c spl2	0.68	0.02	0.66 , 0.71	1.08	0.03	1.03 , 1.14	1.02	0.02	0.99 , 1.06									
Previous HbA1c spl3	1.4	0.02	1.35 , 1.45	1.16	0.02	1.12 , 1.21	1.02	0.02	0.98 , 1.05									
BMI in previous interval	1.03	0.01	1.02 , 1.04	0.88	0.01	0.86 , 0.9	0.98	0.01	0.96 , 0.99									

Table 15.4 Parameter estimates from multinomial logistic regression model for probability of censoring, denominator model for weights (LEFT) and numerator model for stabilisation (RIGHT) Covariate specification C

	DENOMINATOR MODEL									NUMERATOR MODEL								
	Censoring due initiation of other antidiabetic medication			Censoring due to death			Censoring due to transfer out			Censoring due initiation of other antidiabetic medication			Censoring due to death			Censoring due to transfer out		
	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI
BASELINE FIXED																		
Time (months) since study entry																		
Time since study entry (months) spl 1	0.91	0.03	0.86 , 0.96	1.42	0.06	1.26 , 1.62	1.06	0.05	0.96 , 1.17	0.56	0.03	0.53 , 0.59	1.57	0.06	1.38 , 1.77	1.00	0.05	0.9 , 1.09
Time since study entry (months) spl 2	1	0	1 , 1.01	0.99	0.01	0.98 , 1.01	1	0.01	0.99 , 1.01	1.05	0	1.04 , 1.05	1	0.01	0.98 , 1.01	1.00	0.01	0.99 , 1.02
Time (months) between diagnosis and study entry																		
Time between diagnosis and study entry spl 1	0.84	0.03	0.79 , 0.89	1.25	0.04	1.16 , 1.34	1.08	0.03	1.01 , 1.16	0.82	0.04	0.76 , 0.89	1.25	0.04	1.16 , 1.34	1.08	0.04	1.01 , 1.16
Time between diagnosis and study entry spl 2	1.03	0.02	0.99 , 1.07	0.93	0.03	0.88 , 0.99	1.03	0.03	0.98 , 1.08	0.84	0.03	0.8 , 0.9	0.94	0.03	0.89 , 1	1.02	0.03	0.97 , 1.07
Age at diagnosis (years)																		
32-44																		
45-49	0.94	0.03	0.89 , 0.99	2.16	0.19	1.49 , 3.13	0.73	0.05	0.66 , 0.8	0.83	0.03	0.79 , 0.87	1.99	0.19	1.38 , 2.92	0.69	0.05	0.63 , 0.76
60-74	0.9	0.03	0.84 , 0.95	5.87	0.19	4.06 , 8.41	0.55	0.06	0.49 , 0.61	0.66	0.03	0.62 , 0.7	5.75	0.18	4.01 , 8.25	0.51	0.05	0.46 , 0.57
75-89	0.77	0.04	0.71 , 0.84	16.3	0.19	11.3 , 23.6	0.76	0.06	0.67 , 0.86	0.56	0.04	0.52 , 0.61	18.4	0.19	12.8 , 26.6	0.73	0.06	0.64 , 0.83
Gender	1.05	0.02	1.02 , 1.09	0.78	0.04	0.72 , 0.84	0.95	0.03	0.9 , 1.02	0.95	0.02	0.92 , 0.99	0.8	0.04	0.74 , 0.87	0.95	0.03	0.89 , 1.01
Smoking Status																		
Non																		
Current	1.05	0.02	1 , 1.09	2.2	0.05	1.99 , 2.46	1.06	0.04	0.98 , 1.15	1.11	0.02	1.05 , 1.15	2.27	0.05	2.05 , 2.53	1.07	0.04	0.98 , 1.16
Ex	1.06	0.02	1.02 , 1.09	1.4	0.04	1.28 , 1.52	0.99	0.03	0.92 , 1.06	1.04	0.02	1 , 1.07	1.39	0.04	1.27 , 1.51	0.98	0.03	0.91 , 1.05
Alcohol consumption																		

non_drinker																		
ex-drinker	1.07	0.04	0.98 , 1.16	1.28	0.08	1.08 , 1.51	0.93	0.07	0.8 , 1.07	1.02	0.04	0.94 , 1.12	1.3	0.08	1.11 , 1.54	0.92	0.07	0.8 , 1.07
current drinker unknown	1.15	0.06	1.02 , 1.3	1.58	0.11	1.26 , 1.97	1.4	0.1	1.17 , 1.7	1.17	0.07	1.04 , 1.34	1.6	0.11	1.27 , 1.99	1.43	0.1	1.2 , 1.73
rare drinker <2u/d	0.96	0.03	0.91 , 1.02	0.99	0.07	0.87 , 1.13	0.85	0.05	0.77 , 0.95	0.95	0.03	0.9 , 1.01	0.97	0.06	0.85 , 1.11	0.85	0.05	0.77 , 0.94
moderate drinker 3-6u/d	0.98	0.03	0.93 , 1.03	0.89	0.06	0.79 , 1	0.84	0.05	0.77 , 0.93	0.96	0.03	0.9 , 1.01	0.87	0.06	0.77 , 0.97	0.84	0.05	0.76 , 0.92
excessive drinker >6u/d	0.96	0.04	0.89 , 1.03	1.35	0.08	1.15 , 1.58	0.87	0.07	0.77 , 0.99	0.89	0.04	0.83 , 0.95	1.32	0.08	1.13 , 1.54	0.85	0.07	0.75 , 0.97
Year of diabetes onset																		
1990-1994																		
1995-2000	0.54	0.18	0.39 , 0.76	1.28	0.22	0.83 , 1.99	0.81	0.24	0.51 , 1.28	0.5	0.16	0.36 , 0.68	1.22	0.23	0.79 , 1.92	0.79	0.24	0.5 , 1.27
2001-2005	0.34	0.18	0.24 , 0.48	1.21	0.22	0.78 , 1.88	0.82	0.23	0.52 , 1.3	0.3	0.16	0.22 , 0.41	1.14	0.23	0.73 , 1.79	0.79	0.24	0.5 , 1.26
2005 onwards	0.27	0.18	0.19 , 0.38	1.14	0.23	0.73 , 1.77	0.86	0.24	0.54 , 1.36	0.22	0.16	0.16 , 0.3	1.07	0.23	0.68 , 1.68	0.84	0.24	0.53 , 1.35
Use of anti HT in year prior to study entry	0.92	0.03	0.88 , 0.97	1.55	0.07	1.35 , 1.8	1.23	0.06	1.11 , 1.38	0.95	0.02	0.92 , 0.99	1.26	0.05	1.14 , 1.38	0.9	0.03	0.84 , 0.96
Use of statin in year prior to study entry	0.93	0.02	0.9 , 0.97	0.95	0.05	0.86 , 1.04	1.12	0.04	1.03 , 1.2	0.94	0.02	0.9 , 0.98	0.8	0.04	0.74 , 0.87	0.98	0.03	0.92 , 1.05
Use of NSAID in year prior to study entry	1.07	0.02	1.02 , 1.12	0.9	0.05	0.81 , 0.99	1.04	0.04	0.96 , 1.13	1.07	0.02	1.03 , 1.12	0.94	0.05	0.85 , 1.03	0.98	0.04	0.9 , 1.05
HbA1c at study entry																		
<6%																		
6% - 6.5%	1.01	0.04	0.92 , 1.09	1.09	0.06	0.97 , 1.23	0.96	0.05	0.86 , 1.06	1.32	0.04	1.22 , 1.42	0.95	0.06	0.85 , 1.06	0.95	0.05	0.86 , 1.05
6.5%-7%	1.11	0.04	1.02 , 1.2	1.15	0.07	1.01 , 1.31	1.16	0.06	1.04 , 1.3	1.72	0.04	1.6 , 1.84	0.99	0.06	0.88 , 1.11	1.15	0.05	1.04 , 1.27
7% - 8%	1.12	0.04	1.03 , 1.21	1.3	0.07	1.13 , 1.49	1.14	0.06	1.02 , 1.28	2.23	0.04	2.08 , 2.41	1.14	0.06	1 , 1.28	1.17	0.05	1.06 , 1.3
8%-10%	1.11	0.04	1.02 , 1.2	1.42	0.08	1.21 , 1.68	1.22	0.07	1.07 , 1.39	2.97	0.04	2.75 , 3.22	1.34	0.08	1.15 , 1.57	1.31	0.06	1.16 , 1.46
>10%	1.23	0.04	1.13 , 1.34	1.28	0.11	1.03 , 1.6	1.08	0.08	0.93 , 1.27	4.57	0.04	4.22 , 4.95	1.21	0.1	0.98 , 1.48	1.21	0.07	1.05 , 1.39
BMI at study entry																		
<25																		
25-29	0.7	0.04	0.64 , 0.76	1.01	0.06	0.9 , 1.14	1.05	0.07	0.92 , 1.2	0.71	0.03	0.66 , 0.76	0.59	0.05	0.53 , 0.65	0.91	0.05	0.83 , 1.01
30-34	0.57	0.05	0.51 , 0.63	1.32	0.09	1.12 , 1.57	1.09	0.08	0.92 , 1.3	0.66	0.03	0.62 , 0.7	0.53	0.06	0.47 , 0.59	0.87	0.05	0.78 , 0.96
35+	0.5	0.06	0.44 , 0.57	1.75	0.12	1.38 , 2.23	1.04	0.11	0.84 , 1.28	0.69	0.03	0.64 , 0.74	0.61	0.06	0.53 , 0.68	0.83	0.06	0.74 , 0.92

History of CVD at study entry	0.99	0.05	0.9 , 1.09	0.74	0.07	0.64 , 0.85	0.83	0.09	0.69 , 0.98	1.16	0.03	1.09 , 1.22	1.77	0.04	1.63 , 1.93	1.08	0.05	0.99 , 1.19
History of CKD at study entry	1.15	0.06	1.03 , 1.28	1.12	0.08	0.96 , 1.31	0.9	0.1	0.76 , 1.09	1.32	0.05	1.2 , 1.45	1.52	0.07	1.32 , 1.73	0.91	0.08	0.78 , 1.06
TIME UPDATED																		
Medication in previous interval	2.44	0.02	2.34 , 2.56	0.89	0.05	0.81 , 0.98	0.88	0.04	0.82 , 0.94	4.26	0.02	4.06 , 4.44	0.83	0.05	0.76 , 0.91	0.85	0.04	0.79 , 0.91
Use of anti HT in previous year	1.07	0.02	1.03 , 1.13	1.12	0.05	1.01 , 1.25	0.86	0.05	0.79 , 0.94									
Use of statin in previous year	1.12	0.02	1.06 , 1.16	0.7	0.05	0.64 , 0.77	0.79	0.04	0.73 , 0.86									
Use of NSAID in previous year	1.16	0.03	1.11 , 1.22	0.68	0.07	0.59 , 0.79	0.66	0.06	0.58 , 0.73									
History of CVD	1.14	0.05	1.04 , 1.25	2.64	0.07	2.29 , 3	1.34	0.08	1.14 , 1.57									
History of CKD	1.31	0.04	1.22 , 1.42	1.49	0.05	1.34 , 1.65	1.02	0.06	0.9 , 1.15									
HbA1c in previous interval																		
<6%																		
6% - 6.5%	1.52	0.08	1.31 , 1.79	0.7	0.06	0.63 , 0.79	1	0.05	0.9 , 1.11									
6.5%-7%	2.75	0.07	2.36 , 3.16	0.66	0.06	0.59 , 0.75	0.93	0.05	0.84 , 1.04									
7% - 8%	11.7	0.07	10.2 , 13.5	0.7	0.07	0.61 , 0.79	1.02	0.06	0.91 , 1.15									
8%-10%	38.4	0.07	33.5 , 44.2	0.94	0.09	0.79 , 1.14	1.17	0.07	1.02 , 1.35									
>10%	66.0	0.08	56.8 , 76.7	0.96	0.17	0.68 , 1.35	1.34	0.1	1.09 , 1.63									
Bmi in previous interval																		
<25																		
25-29	0.93	0.04	0.86 , 1.01	0.47	0.06	0.41 , 0.52	0.84	0.06	0.74 , 0.95									
30-34	1	0.05	0.91 , 1.11	0.32	0.09	0.27 , 0.39	0.77	0.08	0.66 , 0.9									
35+	1.2	0.06	1.06 , 1.34	0.31	0.13	0.24 , 0.4	0.8	0.1	0.66 , 0.98									

APPENDIX 16 SENSITIVITY ANALYSIS RESULTS FOR METFORMIN AND CANCER

Table 16.1 Sensitivity analysis 1: Hazard ratios (HRs) (approximated from a pooled logistic regression) for metformin vs diet only on risk of cancer obtained from models assuming cancer not present until diagnosis.

	MVRS (A)			SIMPLIFIED (B)			CATEGORISED (C)		
	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value
Model 1 - basic baseline adjustment	0.91	(0.84 , 0.99)	0.023	0.90	(0.83 , 0.98)	0.017	0.88	(0.81 , 0.96)	0.003
Model 2 - Full baseline adjustment	0.90	(0.82 , 0.99)	0.027	0.90	(0.82 , 0.99)	0.025	0.88	(0.8 , 0.97)	0.008
Model 3 - Baseline & time updated adjustment	0.89	(0.81 , 0.98)	0.019	0.89	(0.81 , 0.98)	0.014	0.88	(0.8 , 0.97)	0.01
Model 4 - IPTW weighted	0.92	(0.8 , 1.05)	0.223	0.91	(0.79 , 1.06)	0.221	0.91	(0.79 , 1.04)	0.161
Model 5 - IPTW and censoring weighted	0.94	(0.81 , 1.08)	0.37	0.93	(0.8 , 1.08)	0.329	0.92	(0.8 , 1.06)	0.255

Table 16.2 Sensitivity analysis 2: Hazard ratios (HRs) (approximated from a pooled logistic regression) for metformin vs diet only on risk of cancer obtained from models assuming cancer present 12 months prior to diagnosis.

	MVRS (A)			SIMPLIFIED (B)			CATEGORISED (C)		
	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value
Model 1 - basic baseline adjustment	0.95	(0.87 , 1.04)	0.232	0.95	(0.87 , 1.04)	0.237	0.92	(0.85 , 1.01)	0.086
Model 2 - Full baseline adjustment	0.98	(0.88 , 1.08)	0.645	0.99	(0.89 , 1.09)	0.799	0.96	(0.87 , 1.06)	0.387
Model 3 - Baseline & time updated adjustment	0.98	(0.89 , 1.09)	0.726	0.99	(0.89 , 1.09)	0.792	0.96	(0.87 , 1.06)	0.44
Model 4 - IPTW weighted	0.98	(0.85 , 1.14)	0.835	1.01	(0.86 , 1.18)	0.923	1.03	(0.89 , 1.19)	0.684
Model 5 - IPTW and censoring weighted	1.01	(0.87 , 1.17)	0.905	1.03	(0.88 , 1.21)	0.712	1.06	(0.92 , 1.23)	0.398

Table 16.3 Sensitivity analysis 3: Hazard ratios (HRs) (approximated from a pooled logistic regression) for metformin vs diet only on risk of cancer where non melanoma skin cancer is excluded from the set of cancers.

	MVRS (A)			SIMPLIFIED (B)			CATEGORISED (C)		
	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value
Model 1 - basic baseline adjustment	0.96	(0.87 , 1.06)	0.424	0.96	(0.87 , 1.06)	0.418	0.94	(0.85 , 1.04)	0.217
Model 2 - Full baseline adjustment	0.97	(0.87 , 1.08)	0.587	0.98	(0.87 , 1.09)	0.69	0.94	(0.84 , 1.06)	0.319
Model 3 - Baseline & time updated adjustment	0.96	(0.85 , 1.08)	0.47	0.97	(0.86 , 1.08)	0.543	0.94	(0.84 , 1.05)	0.262
Model 4 - IPTW weighted	1.00	(0.84 , 1.18)	0.991	0.99	(0.83 , 1.17)	0.883	1.02	(0.86 , 1.2)	0.823
Model 5 - IPTW and censoring weighted	1.03	(0.86 , 1.22)	0.772	1.03	(0.86 , 1.22)	0.781	1.05	(0.89 , 1.25)	0.553

Table 16.4 Sensitivity analysis 4: Hazard ratios (HRs) (approximated from a pooled logistic regression) for metformin vs diet only on risk of cancer obtained from models allowing baseline covariates for patients initiating treatment at time of diagnosis to occur after date of first treatment (but before study entry).

	MVRS (A)			SIMPLIFIED (B)			CATEGORISED (C)		
	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value
Model 1 - basic baseline adjustment	0.92	(0.84 , 1)	0.039	0.92	(0.84 , 1)	0.038	0.89	(0.82 , 0.97)	0.007
Model 2 - Full baseline adjustment	0.95	(0.86 , 1.04)	0.287	0.96	(0.87 , 1.05)	0.371	0.93	(0.85 , 1.02)	0.119
Model 3 - Baseline & time updated adjustment	0.94	(0.85 , 1.03)	0.182	0.94	(0.86 , 1.04)	0.214	0.92	(0.83 , 1.01)	0.073
Model 4 - IPTW weighted	0.94	(0.81 , 1.08)	0.363	0.95	(0.82 , 1.1)	0.503	0.97	(0.85 , 1.11)	0.654
Model 5 - IPTW and censoring weighted	0.95	(0.82 , 1.1)	0.51	0.97	(0.84 , 1.13)	0.698	0.99	(0.86 , 1.14)	0.925

Table 16.5 Sensitivity analysis 5: Hazard ratios (HRs) (approximated from a pooled logistic regression) for metformin vs diet only on risk of cancer obtained from models allowing use of baseline covariates for patients initiating treatment in same interval that they get complete data.

	MVRS (A)			SIMPLIFIED (B)			CATEGORISED (C)		
	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value
Model 1 - basic baseline adjustment	0.92	(0.85 , 1)	0.051	0.92	(0.85 , 1)	0.049	0.90	(0.83 , 0.97)	0.01
Model 2 - Full baseline adjustment	0.96	(0.87 , 1.05)	0.346	0.96	(0.88 , 1.06)	0.427	0.94	(0.85 , 1.03)	0.156
Model 3 - Baseline & time updated adjustment	0.95	(0.86 , 1.04)	0.253	0.95	(0.86 , 1.04)	0.279	0.93	(0.85 , 1.02)	0.117
Model 4 - IPTW weighted	0.94	(0.82 , 1.08)	0.377	0.95	(0.82 , 1.11)	0.535	0.97	(0.85 , 1.11)	0.695
Model 5 - IPTW and censoring weighted	0.96	(0.83 , 1.1)	0.544	0.97	(0.84 , 1.13)	0.738	1.00	(0.87 , 1.15)	0.988

Table 16.6 Sensitivity analysis 6: Hazard ratios (HRs) (approximated from a pooled logistic regression) for metformin vs diet only on risk of cancer obtained from models using covariate information from the same interval to predict treatment

	MVRS (A)			SIMPLIFIED (B)			CATEGORISED (C)		
	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value
Model 1 - basic baseline adjustment	0.93	(0.85 , 1.01)	0.088	0.93	(0.85 , 1.01)	0.087	0.90	(0.83 , 0.98)	0.021
Model 2 - Full baseline adjustment	0.96	(0.87 , 1.06)	0.396	0.97	(0.88 , 1.06)	0.488	0.94	(0.85 , 1.03)	0.183
Model 3 - Baseline & time updated adjustment	0.95	(0.86 , 1.05)	0.306	0.95	(0.86 , 1.05)	0.332	0.93	(0.84 , 1.03)	0.152
Model 4 - IPTW weighted	0.97	(0.82 , 1.14)	0.707	1.00	(0.84 , 1.18)	0.964	1.01	(0.86 , 1.18)	0.927
Model 5 - IPTW and censoring weighted	0.99	(0.84 , 1.18)	0.945	1.02	(0.86 , 1.22)	0.801	1.04	(0.89 , 1.23)	0.615

Table 16.7 Sensitivity analysis 7: Hazard ratios (HRs) (approximated from a pooled logistic regression) for metformin vs diet only on risk of cancer obtained from models using 3 months instead of 1 month as the interval in which to predict treatment

	MVRS (A)			SIMPLIFIED (B)			CATEGORISED (C)		
	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value
Model 1 - basic baseline adjustment	0.92	(0.84 , 1.01)	0.068	0.92	(0.84 , 1.01)	0.073	0.90	(0.82 , 0.98)	0.018
Model 2 - Full baseline adjustment	0.95	(0.86 , 1.04)	0.271	0.96	(0.87 , 1.06)	0.393	0.93	(0.84 , 1.03)	0.143
Model 3 - Baseline & time updated adjustment	0.95	(0.86 , 1.05)	0.292	0.95	(0.86 , 1.05)	0.316	0.93	(0.84 , 1.03)	0.146
Model 4 - IPTW weighted	0.95	(0.83 , 1.08)	0.413	0.96	(0.86 , 1.06)	0.414	0.98	(0.86 , 1.11)	0.702
Model 5 - IPTW and censoring weighted	0.97	(0.85 , 1.11)	0.677	0.97	(0.86 , 1.08)	0.557	1.01	(0.88 , 1.15)	0.904

Table 16.8 Sensitivity analysis 8: Hazard ratios (HRs) (approximated from a pooled logistic regression) for metformin vs diet only on risk of cancer obtained from models where a patients exposure status changes 1 year after the first metformin prescription.

	MVRS (A)			SIMPLIFIED (B)			CATEGORISED (C)		
	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value
Model 1 - basic baseline adjustment	0.94	(0.86 , 1.04)	0.246	0.94	(0.86 , 1.04)	0.247	0.92	(0.84 , 1.02)	0.103
Model 2 - Full baseline adjustment	0.98	(0.88 , 1.09)	0.664	0.98	(0.89 , 1.09)	0.764	0.96	(0.86 , 1.06)	0.425
Model 3 - Baseline & time updated adjustment	0.96	(0.87 , 1.07)	0.511	0.97	(0.87 , 1.08)	0.595	0.95	(0.85 , 1.05)	0.314
Model 4 - IPTW weighted	1.01	(0.86 , 1.18)	0.949	1.02	(0.86 , 1.21)	0.799	1.03	(0.88 , 1.21)	0.711
Model 5 - IPTW and censoring weighted	1.03	(0.87 , 1.21)	0.743	1.05	(0.88 , 1.25)	0.582	1.07	(0.9 , 1.26)	0.433

Table 16.9 Sensitivity analysis 9: Hazard ratios (HRs) (approximated from a pooled logistic regression) for metformin vs diet only on risk of cancer obtained from models where IPTW and IPCW were estimated separately by calendar periods of 1990-1995, 1995-2000, 2000-2005 and 2005 onwards. (model then restricted to 2000 onwards due to sparsity in earlier calendar periods making weight estimation unreliable).

	MVRS (A)			SIMPLIFIED (B)			CATEGORISED (C)		
	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value	HR	95% Confidence Interval	P value
Model 1 - basic baseline adjustment	0.96	(0.88 , 1.05)	0.428	0.96	(0.88 , 1.05)	0.425	0.94	(0.86 , 1.03)	0.175
Model 2 - Full baseline adjustment	0.99	(0.9 , 1.1)	0.877	1.00	(0.91 , 1.11)	0.987	0.97	(0.88 , 1.08)	0.604
Model 3 - Baseline & time updated adjustment	0.99	(0.89 , 1.09)	0.779	0.99	(0.9 , 1.1)	0.85	0.97	(0.88 , 1.07)	0.533
Model 4 - IPTW weighted	1.00	(0.86 , 1.16)	0.993	1.02	(0.87 , 1.19)	0.845	1.03	(0.89 , 1.19)	0.683
Model 5 - IPTW and censoring weighted	1.00	(0.86 , 1.17)	0.954	1.03	(0.88 , 1.2)	0.731	1.04	(0.89 , 1.21)	0.605

APPENDIX 17 TREATMENT MODELS BY CALENDAR PERIOD

Table 17.1 Estimated OR, standard error and 95% CI for probability of treatment with metformin (denominator models for the IPTW).model fitted separately by calendar time period. Periods pre 1995 and 1995-2000 had insufficient numbers to fit model. Some example differences are highlighted with boxing.

	2000-2005			2005 onwards		
	OR	SE	95% CI	OR	SE	95% CI
Time (months) since study entry						
Time since study entry (months) spl 1	0.99	0.001	0.98 , 0.99	0.98	0.001	0.97 , 0.98
Time since study entry (months) spl 2	1.03	0.007	1.01 , 1.04	1.08	0.008	1.07 , 1.1
Time (months) between diagnosis and study entry						
Time between diagnosis and study entry spl 1	0.85	0.048	0.77 , 0.95	0.72	0.040	0.65 , 0.81
Time between diagnosis and study entry spl 2	0.97	0.038	0.9 , 1.05	1.08	0.040	1 , 1.16
Age at diagnosis (years)						
32-44	1 (ref)			1 (ref)		
45-49	1.13	0.072	1.00 , 1.28	0.92	0.036	0.85 , 0.99
60-74	0.98	0.062	0.87 , 1.11	0.86	0.034	0.79 , 0.93
75-89	0.64	0.048	0.56 , 0.74	0.65	0.031	0.59 , 0.71
Gender	1.14	0.034	1.08 , 1.21	1.13	0.024	1.08 , 1.18
Smoking Status						
Non	1 (ref)			1 (ref)		
Current	0.94	0.038	0.87 , 1.02	1.08	0.032	1.02 , 1.14
Ex	1.04	0.032	0.98 , 1.1	1.09	0.024	1.04 , 1.13
Alcohol consumption						
non_drinker						
ex-drinker	0.98	0.094	0.81 , 1.18	0.95	0.044	0.87 , 1.04
current drinker unknown	0.84	0.086	0.69 , 1.03	0.83	0.084	0.68 , 1.02
rare drinker <2u/d	1.05	0.055	0.94 , 1.16	0.95	0.034	0.89 , 1.02
moderate drinker 3-6u/d	1.04	0.050	0.94 , 1.14	0.94	0.032	0.88 , 1.01
excessive drinker >6u/d	0.96	0.062	0.85 , 1.1	0.91	0.040	0.84 , 1
Use of anti HT in year prior to study entry	0.88	0.042	0.8 , 0.97	0.88	0.039	0.8 , 0.96
Use of statin in year prior to study entry	0.9	0.034	0.83 , 0.97	0.88	0.027	0.82 , 0.93
Use of NSAID in year prior to study entry	1.14	0.042	1.06 , 1.22	1.05	0.032	0.99 , 1.11
HbA1c at study entry						
<6%	1 (ref)			1 (ref)		
6% - 6.5% base	0.87	0.044	0.79 , 0.96	0.89	0.051	0.79 , 1
6.5%-7% base	0.79	0.042	0.71 , 0.88	0.78	0.046	0.69 , 0.87
7% - 8% base	0.67	0.037	0.6 , 0.74	0.59	0.037	0.52 , 0.67
8%-10% base	0.50	0.035	0.44 , 0.58	0.39	0.028	0.34 , 0.45
>10% base	0.47	0.042	0.4 , 0.56	0.39	0.033	0.33 , 0.46
BMI at study entry						
<25	1 (ref)			1 (ref)		
25-29 base	1.20	0.088	1.04 , 1.39	0.85	0.059	0.75 , 0.98
30-34 base	1.09	0.096	0.92 , 1.3	0.82	0.065	0.7 , 0.96
35+ base	0.91	0.099	0.73 , 1.12	0.72	0.067	0.6 , 0.87
History of CVD at study entry	0.95	0.087	0.8 , 1.14	0.85	0.072	0.72 , 1.00
History of CKD at study entry	0.47	0.160	0.24 , 0.92	0.85	0.057	0.74 , 0.97
Use of anti HT in previous year	1.03	0.038	0.95 , 1.1	1.10	0.033	1.04 , 1.17
Use of statin in previous year	1.42	0.054	1.32 , 1.53	1.64	0.053	1.53 , 1.74
Use of NSAID in previous year	1.21	0.061	1.1 , 1.33	1.16	0.052	1.06 , 1.26
History of CVD	0.98	0.085	0.83 , 1.16	1.11	0.089	0.95 , 1.3
History of CKD	0.89	0.057	0.78 , 1.01	0.86	0.048	0.77 , 0.95
HbA1c in previous interval						
<6%	1 (ref)			1 (ref)		
6% - 6.5%	2.35	0.234	1.93 , 2.86	2.35	0.195	2 , 2.77
6.5%-7%	7.26	0.682	6.04 , 8.73	7.38	0.596	6.3 , 8.65
7% - 8%	38.97	3.567	32.57 , 46.63	47.11	3.801	40.22 , 55.18
8%-10%	159.3	15.6	131.49 , 192.93	233.7	20.24	197.25 , 276.96
>10%	278.3	34.1	218.88 , 353.85	378.5	38.0	310.96 , 460.8
BMI in previous interval						
<25	1 (ref)			1 (ref)		
25-29	1.17	0.083	1.02 , 1.34	1.49	0.098	1.32 , 1.7
30-34	1.51	0.129	1.28 , 1.78	1.68	0.129	1.44 , 1.95
35+	2.00	0.213	1.62 , 2.46	2.17	0.196	1.82 , 2.59

APPENDIX 18 DESCRIPTIVE RESULTS FOR SULFONYLUREA VS DIET ONLY

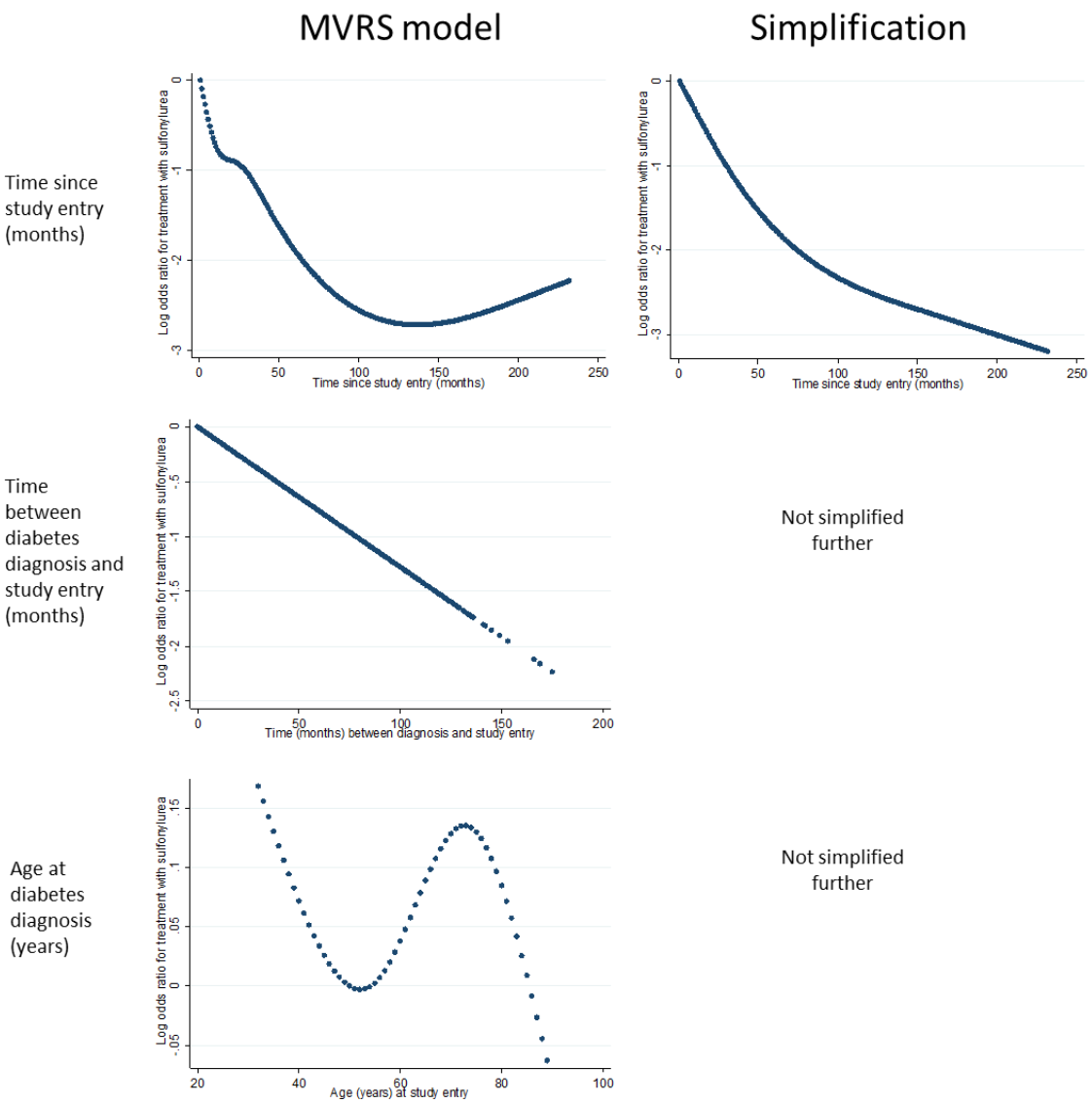
Baseline demographics

Table 18.1 Baseline demographics of patients eligible for sulfonylurea study entry, by medication choice (sulfonylurea or diet) at time of study entry.

	No Therapy N = 49,524	**Sulfonylurea N=962**	**Total N=50486**
Mean (SD) median, 25th %ile – 75th %ile			
Age at diagnosis (years)	62.2 (12) 63 ,54 - 71	59.8 (13.4) 60 ,49 - 70	62.2 (12) 63 ,54 - 71
Time (years) to complete data	3.8 (10.2) 1 ,0 - 3	0 (0) 0 ,0 - 0	3.8 (10.1) 1 ,0 - 3
HbA1c (%)	7.2 (1.6) 6.8 ,6.2 - 7.7	11 (2.5) 11 ,9 - 12.7	7.3 (1.7) 6.8 ,6.3 - 7.7
BMI (kg/m^2)	31.6 (6.3) 30.7 ,27.3 - 34.9	27.3 (5.7) 26.1 ,23.8 - 30	31.5 (6.3) 30.6 ,27.2 - 34.9
N (%)			
Gender			
Male	27763 (56.1)	594 (61.8)	28357 (56.2)
Female	21761 (43.9)	368 (38.3)	22129 (43.8)
History of Chronic Kidney Disease			
No	46463 (93.8)	892 (92.7)	47355 (93.8)
Yes	3061 (6.2)	70 (7.3)	3131 (6.2)
History of Cardiovascular Disease			
No	41868 (84.5)	856 (89)	42724 (84.6)
Yes	7656 (15.5)	106 (11)	7762 (15.4)
Use of statins in previous year			
No	25035 (50.6)	589 (61.2)	25624 (50.8)
Yes	24489 (49.5)	373 (38.8)	24862 (49.3)
Use of NSAID in previous year			
No	39575 (79.9)	793 (82.4)	40368 (80)
Yes	9949 (20.1)	169 (17.6)	10118 (20)
Use of Anti HT in previous year			
No	18048 (36.4)	519 (54)	18567 (36.8)
Yes	31476 (63.6)	443 (46.1)	31919 (63.2)
Smoking Status			
Non	20132 (40.7)	403 (41.9)	20535 (40.7)
Current	8746 (17.7)	240 (25)	8986 (17.8)
Ex	20646 (41.7)	319 (33.2)	20965 (41.5)
Alcohol consumption			
non-drinker	5770 (11.7)	105 (10.9)	5875 (11.6)
ex-drinker	3474 (7)	74 (7.7)	3548 (7)
current drinker quantity unknown	979 (2)	24 (2.5)	1003 (2)
rare drinker <2u/d	11543 (23.3)	202 (21)	11745 (23.3)
moderate drinker 3-6u	22934 (46.3)	461 (47.9)	23395 (46.3)
excessive drinker >6u	4824 (9.7)	96 (10)	4920 (9.8)
Hypoglycaemia in past 3 months			
No	49486 (99.9)	962 (100)	50448 (99.9)
Yes	38 (0.1)	0 (0)	38 (0.1)
Proteinuria in past 3 months			
No	49424 (99.8)	962 (100)	50386 (99.8)
Yes	100 (0.2)	0 (0)	100 (0.2)
Oedema in past 3 months			
No	48550 (98)	945 (98.2)	49495 (98)
Yes	974 (2)	17 (1.8)	991 (2)
Anaemia in past 3 months			
No	49284 (99.5)	960 (99.8)	50244 (99.5)
Yes	240 (0.5)	2 (0.2)	242 (0.5)
Calendar Year of onset			
1990 - 1995	134 (0.3)	0 (0)	134 (0.3)
1995 - 2000	1708 (3.5)	34 (3.5)	1742 (3.5)
2000-2005	12764 (25.8)	295 (30.7)	13059 (25.9)
post 2005	34918 (70.5)	633 (65.8)	35551 (70.4)

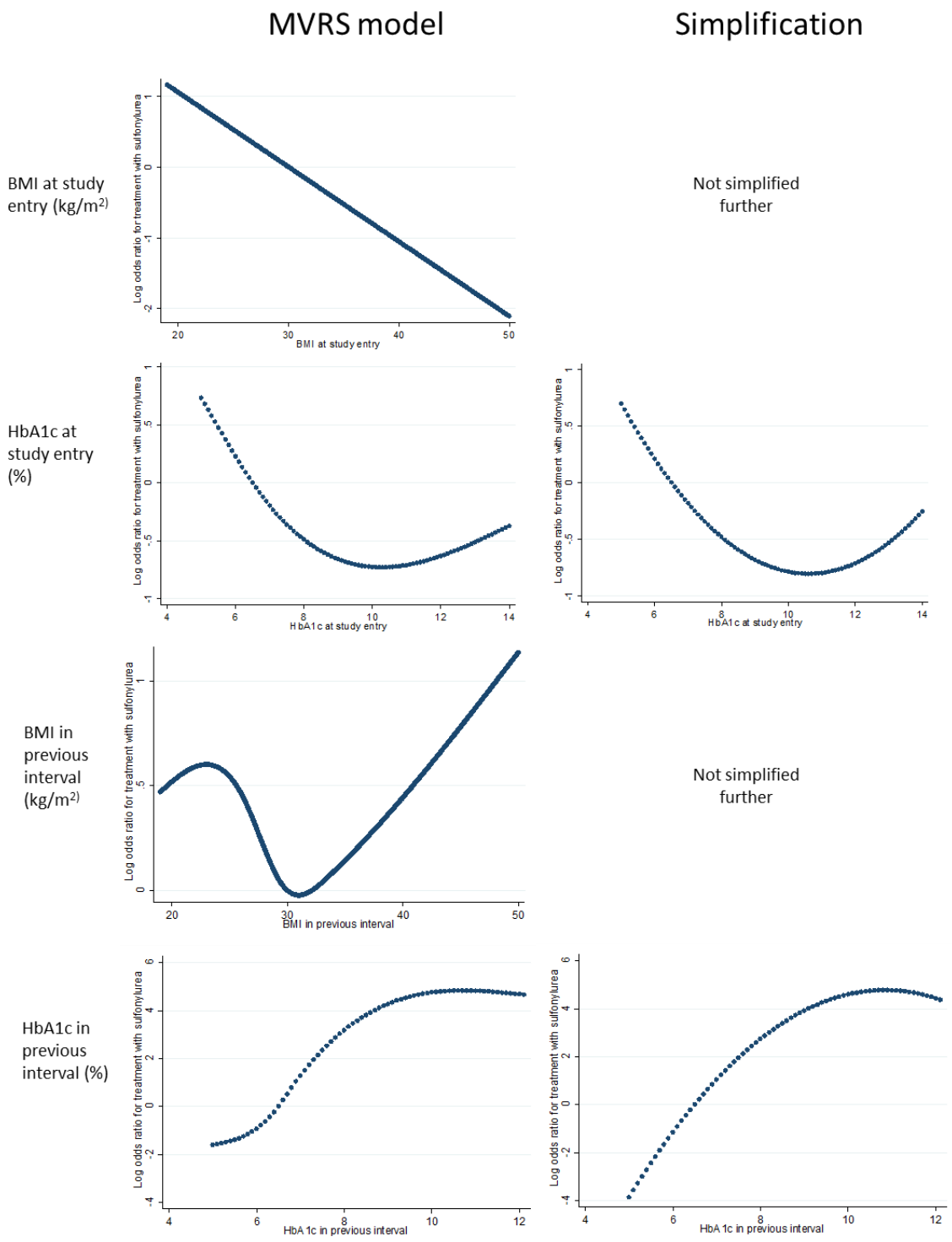
Estimated spline associations

Figure 18.1 Associations between continuous variables and treatment with sulfonylurea from multivariable model where spline knots and knot points decided by iterative function MVRS (left). If simplification deemed possible, this is shown on the right.



Reference time = 1, reference age = 50.

Figure 18.2 Associations between continuous variables and treatment with sulfonylurea from multivariable model where spline knots and knot points decided by iterative function MVRS (left) . If simplification deemed possible, this is shown on the right.



Reference BMI = 30 kg/m^2, Reference HbA1c = 6.5%

Distributions of IPTW and IPCW

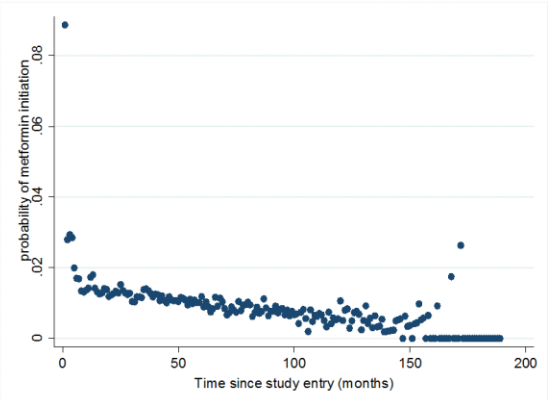
Table 18.2 Distribution of inverse probability of treatment weights for sulfonylurea vs diet analysis

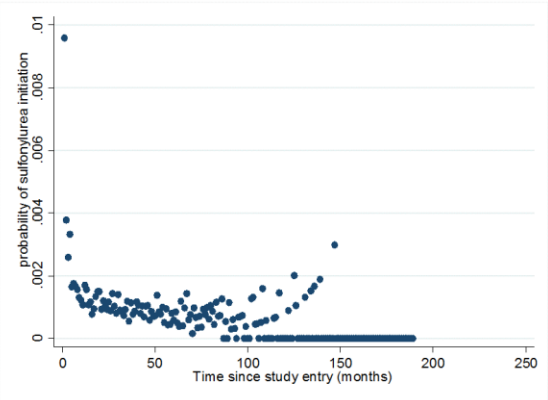
Specification	Truncation	Mean	SD	1 %ile	25 %ile	50 %ile	75 %ile	99%ile	Min	Max
A	None	29.244	432.94	1	1.004	1.012	1.041	488.48	1	70273.3
	1% & 99%	1.031	2.67	0.17	0.98	0.99	1	1.91	0.002	1618.8
	10 & 0.1	0.981	0.17	0.17	0.98	0.99	1	1.91	0.17	1.9
B	None	1.003	0.44	0.17	0.98	0.99	1	1.91	0.1	10
	1% & 99%	37.097	810.81	1	1	1.01	1.04	451.01	1	99425.9
	10 & 0.1	1.04	2.65	0.19	0.98	0.99	1	1.82	0.002	1082.8
C	None	0.977	0.16	0.19	0.98	0.99	1	1.82	0.187	1.8
	1% & 99%	1.001	0.46	0.19	0.98	0.99	1	1.82	0.1	10
	10 & 0.1	28.858	378.09	1	1	1.01	1.04	517.68	1	36567.8

Table 18.3 Distribution of joint inverse probability of treatment and inverse probability of censoring weights for sulfonylurea vs diet analysis

Specification	Truncation	Mean	SD	1 %ile	25 %ile	50 %ile	75 %ile	99%ile	Min	Max
A	None	1139.98	377202	0.08	0.78	0.9	0.97	5.19	0	251000000
	1% & 99%	0.96	0.61	0.08	0.78	0.9	0.97	5.19	0.08	5.19
	10 & 0.1	1	0.9	0.1	0.78	0.9	0.97	5.19	0.1	10
B	None	4.18	727.47	0.08	0.77	0.9	0.98	5.02	0	417921.5
	1% & 99%	0.96	0.6	0.08	0.77	0.9	0.98	5.02	0.08	5.02
	10 & 0.1	1	0.89	0.1	0.77	0.9	0.98	5.02	0.1	10
C	None	1129.61	375459	0.08	0.78	0.9	0.97	5.38	0	250000000
	1% & 99%	0.96	0.62	0.08	0.78	0.9	0.97	5.38	0.08	5.38
	10 & 0.1	0.99	0.91	0.1	0.78	0.9	0.97	5.38	0.1	10

APPENDIX 19 INVESTIGATIONS INTO APPROPRIATE FOLLOW UP LENGTH





Month	0	50	100	120	150
N untreated	49,625	11,120	2,515	1,223	274

In the sulfonylurea users (right hand plot), the probability of treatment is always low but by 100 months, zero probabilities begin to occur. Therefore the weighting for initiation of sulfonylurea users by this time may become unreliable. Since the strict violation of the positivity assumption does not occur until 150 months, 120 months (10 years) was considered a reasonable cut off.

APPENDIX 20 SPLINE PARAMETERISATIONS AND CATEGORISATIONS USED IN CHAPTER 8

	A (MVRS output) Natural cubic spline with knots at…	**B Categorised**
Time since study entry (months)	1, 10 , 24 , 48, 120	Natural cubic spline with knots at 10, 24 , 48 .
Time between diagnosis and study entry (months)	0 , 1, 3 ,100	as A
HbA1c (%) in interval -1	5, 6.4 , 6.8, 14	<6, [6-6.5), [6.5-7) , [7-8) , [8-10) , 10 +
HbA1c (%) in interval -2	5 , 6, 6.4 , 6.8 , 14	<6, [6-6.5), [6.5-7) , [7-8) , [8-10) , 10 +
BMI in previous interval	20 , 26 , 30 , 33 , 50	<25 , [25-30), [30-35), 35+
SBP in previous interval	100 , 126 , 142 , 200	[100-129] , [130-139] , [140-149], 150+
Age at diagnosis	32 , 56 , 72 , 90	32-44 , 45--59 , 60-74 , 75+
HbA1c (%) at study entry	5 , 6 , 6.4 , 6.9 , 14	<6, [6-6.5), [6.5-7) , [7-8) , [8-10) , 10 +
BMI at study entry	MI, stroke and HbA1c: Linear term ACM: 20, 27, 30, 50	<25 , [25-30), [30-35), 35+
SBP at study entry	100, 139, 200	[100-129] , [130-139] , [140-149], 150+

APPENDIX 21 COMPARISONS OF SPLINE PARAMETERISATIONS FOR ASSOCIATION BETWEEN BASELINE COVARIATES AND RISK OF MI, STROKE, ACM AND HBA1C BETWEEN A GENERAL PARAMETERISATION (LEFT) AND PARAMETERISATION ESTIMATED IN TREATMENT MODEL FITTING STAGE (RIGHT).

Figure 21.1 MI

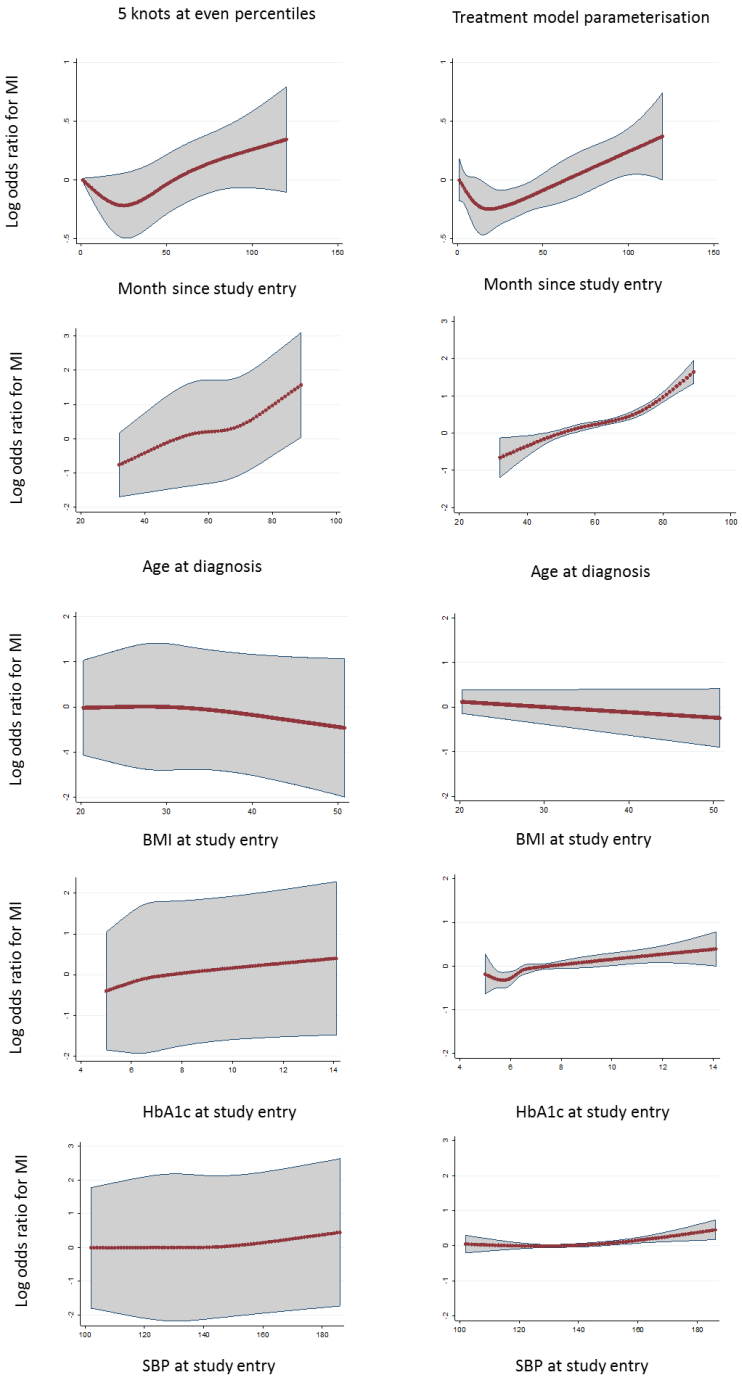


Figure 21.2 Stroke

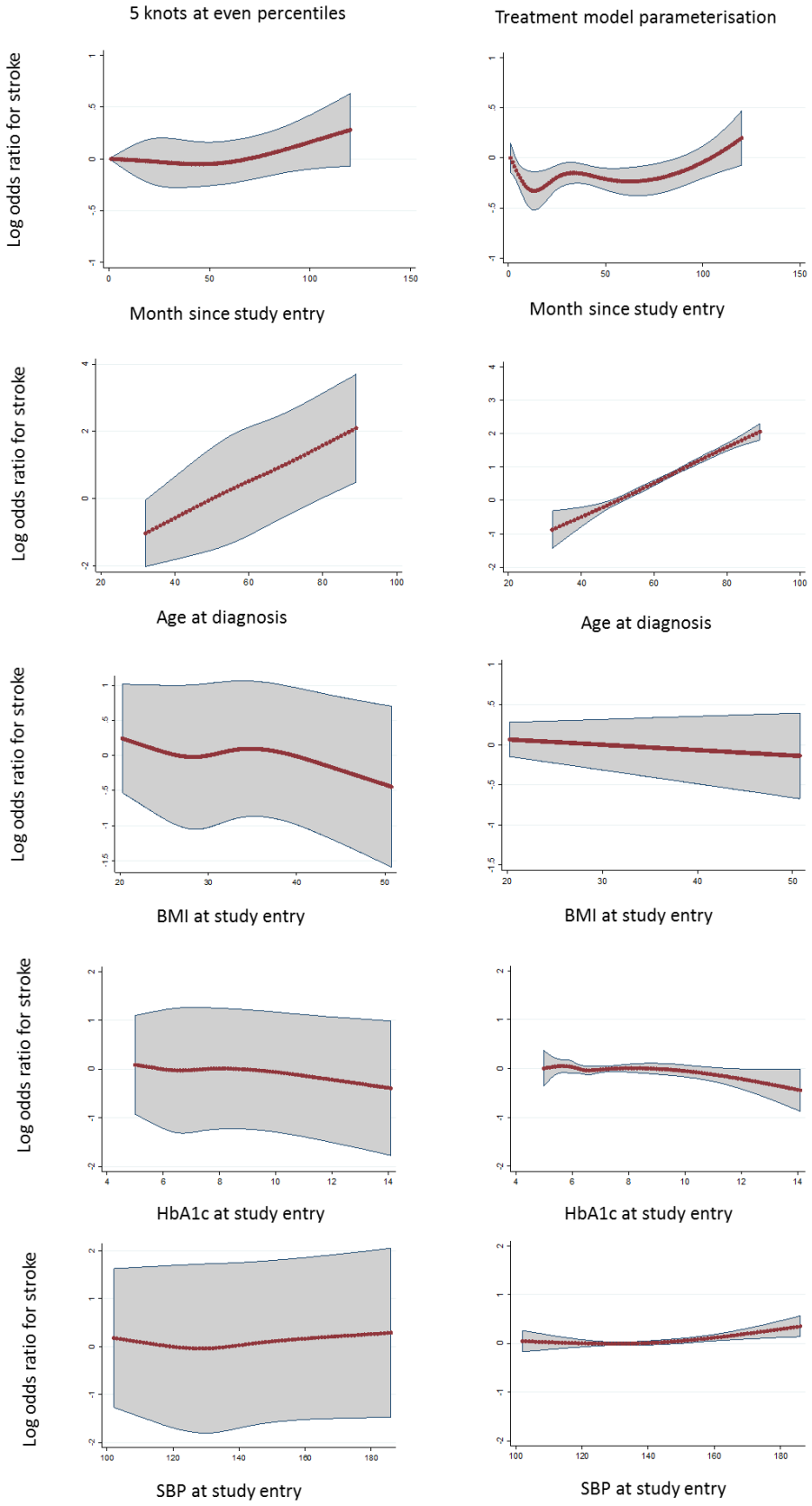


Figure 21.3 All-cause mortality

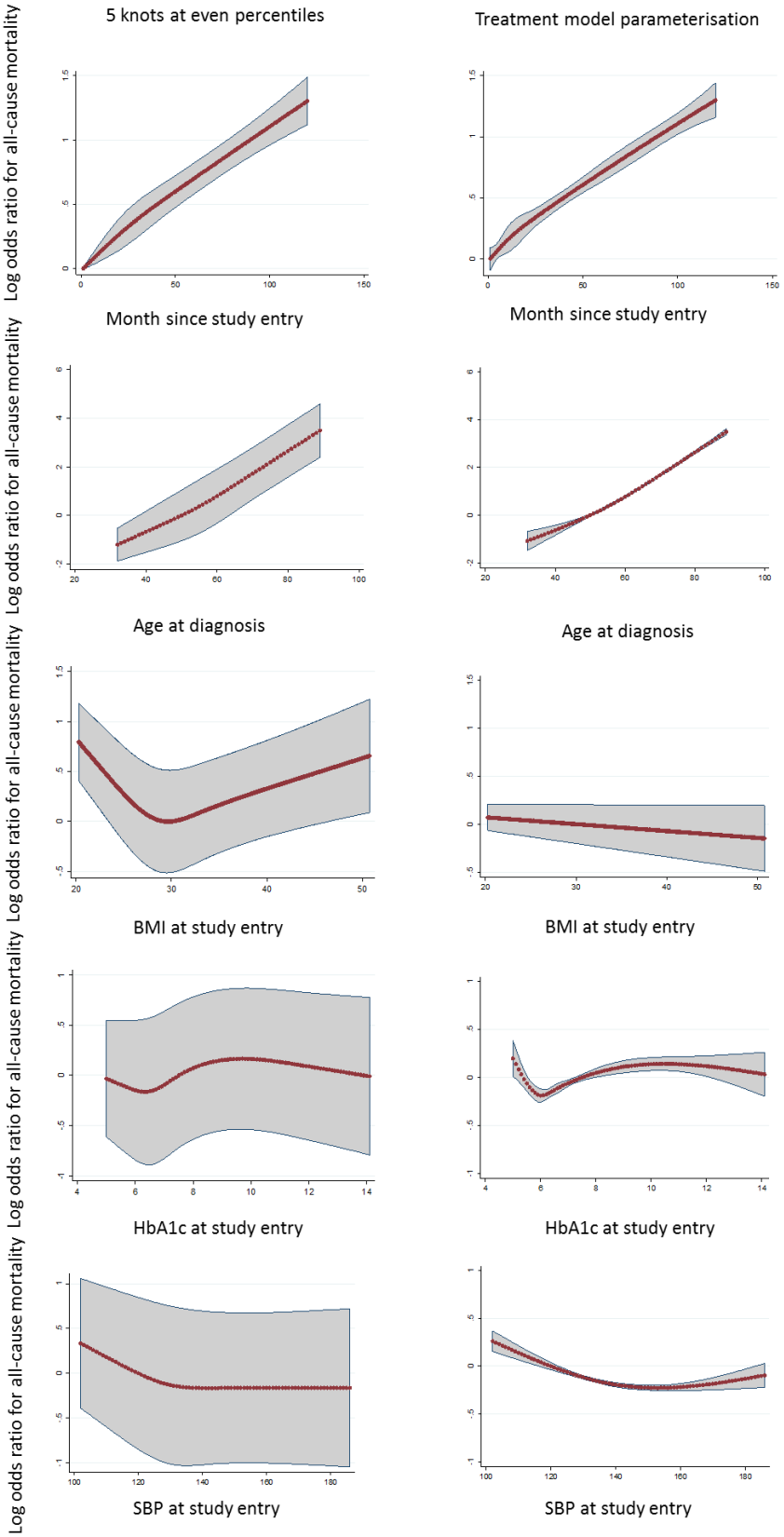


Figure 21.4 Comparison of parameterisations for baseline BMI when looking at risk of all-cause mortality. Right hand pane is parameterisation used in models for censoring and in mortality outcome model as parameterisation from treatment model (centre) deemed unsuitable

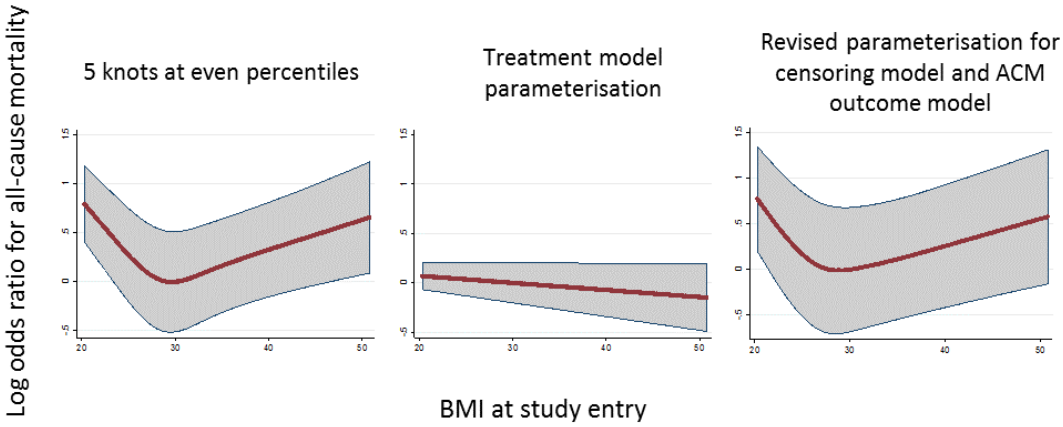
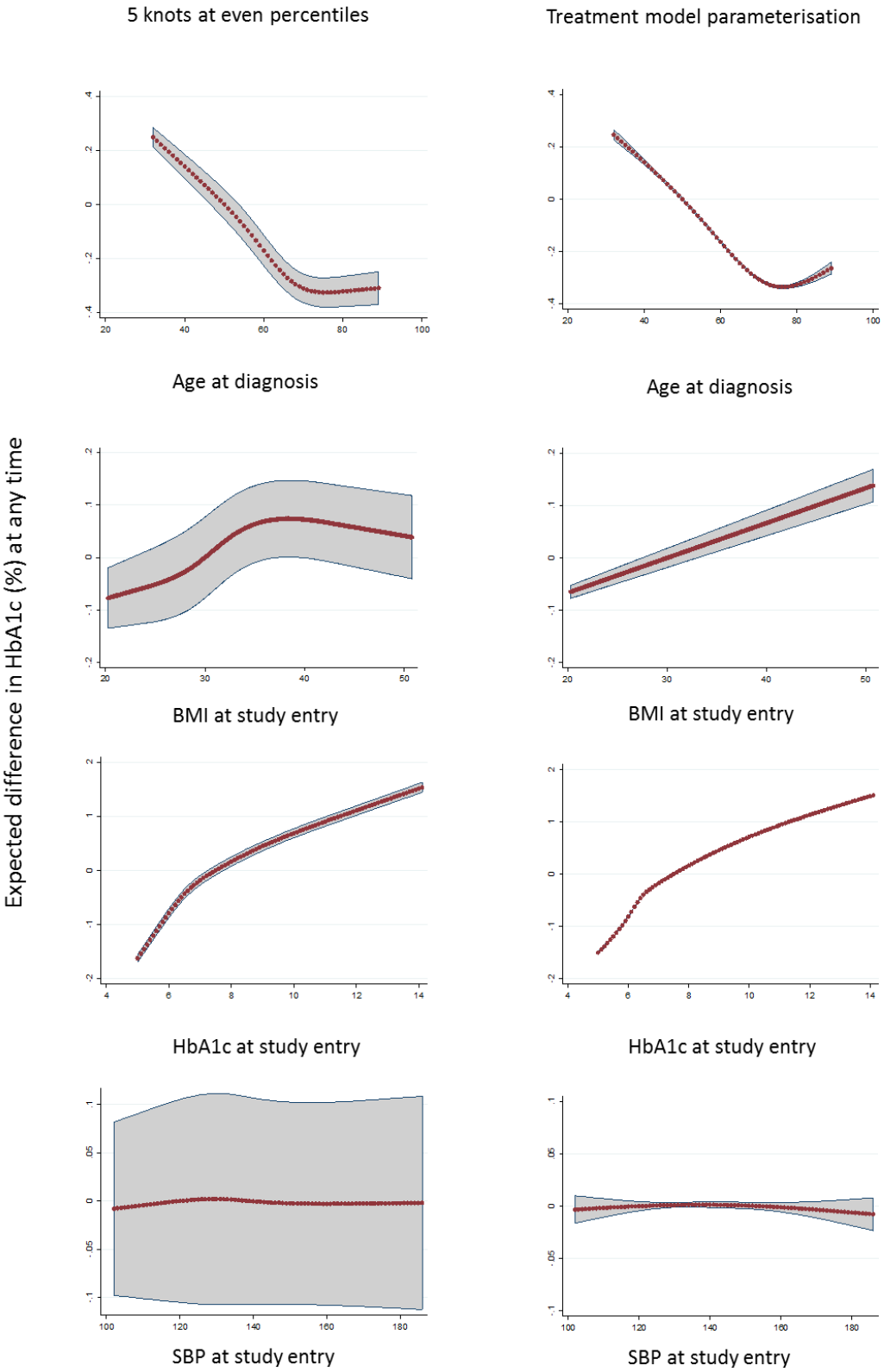


Figure 21.5 HbA1c (repeated measures continuous outcome)



APPENDIX 22 EXAMPLE TREATMENT AND CENSORING MODEL OUTPUTS: FOLLOW UP TO ALL-CAUSE MORTALITY (CATEGORICAL SPECIFICATION).

Treatment model: Log odds ratios for treatment with metformin (Left half) or sulfonylureas (right half) estimated from a multinomial logistic regression model, fitted separately by calendar period.

Table 22.1 2000-2005

	Metformin						Sulfonylurea					
	DENOMINATOR MODEL			**NUMERATOR MODEL**			**DENOMINATOR MODEL**			**NUMERATOR MODEL**		
	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI
Time (months) since study entry												
Time since study entry (months) spl 1	-0.02	0.002	-0.03 , -0.02	-0.03	0.001	-0.03 , -0.02	-0.04	0.004	-0.05 , -0.03	-0.05	0.004	-0.06 , -0.05
Time since study entry (months) spl 2	0.02	0.002	0.01 , 0.02	0.02	0.002	0.02 , 0.03	0.02	0.005	0.01 , 0.03	0.04	0.004	0.03 , 0.05
Time (months) between diagnosis and study entry	-0.01	0.002	-0.01 , 0	0	0.001	0 , 0	-0.02	0.004	-0.03 , -0.01	-0.01	0.003	-0.02 , 0
Age at diagnosis (years)												
32-44												
45-49	0.07	0.060	-0.05 , 0.18	0.05	0.056	-0.06 , 0.16	-0.19	0.136	-0.46 , 0.07	-0.22	0.135	-0.49 , 0.04
60-74	-0.09	0.060	-0.21 , 0.02	-0.22	0.057	-0.34 , -0.11	-0.07	0.138	-0.34 , 0.2	-0.15	0.137	-0.42 , 0.12
75-89	-0.47	0.070	-0.61 , -0.33	-0.72	0.068	-0.86 , -0.59	-0.13	0.154	-0.43 , 0.17	-0.18	0.151	-0.48 , 0.11
Gender	0.13	0.028	0.08 , 0.19	0.13	0.029	0.07 , 0.18	0.02	0.065	-0.11 , 0.14	0	0.067	-0.14 , 0.13
Smoking Status												
Non												
Current	-0.06	0.038	-0.14 , 0.01	0.02	0.038	-0.05 , 0.1	0.04	0.081	-0.12 , 0.2	0.14	0.083	-0.02 , 0.3
Ex	0.04	0.030	-0.02 , 0.1	0.06	0.031	0 , 0.12	0.07	0.068	-0.07 , 0.2	0.09	0.070	-0.05 , 0.22
Alcohol consumption												
non_drinker												
ex-drinker	-0.06	0.086	-0.23 , 0.11	-0.05	0.083	-0.22 , 0.11	-0.09	0.162	-0.41 , 0.23	-0.13	0.163	-0.45 , 0.19
current drinker unknown	-0.13	0.094	-0.31 , 0.06	-0.06	0.090	-0.24 , 0.12	0.01	0.171	-0.33 , 0.34	0.08	0.168	-0.25 , 0.41
rare drinker <2u/d	0.05	0.050	-0.05 , 0.15	0.04	0.050	-0.06 , 0.14	-0.23	0.104	-0.43 , -0.03	-0.26	0.105	-0.47 , -0.05
moderate drinker 3-6u/d	0.01	0.045	-0.08 , 0.1	-0.01	0.045	-0.09 , 0.08	-0.16	0.090	-0.34 , 0.01	-0.23	0.090	-0.41 , -0.05
excessive drinker >6u/d	-0.05	0.062	-0.17 , 0.07	-0.07	0.062	-0.2 , 0.05	-0.28	0.136	-0.54 , -0.01	-0.36	0.141	-0.64 , -0.09
Use of anti HT in year prior to study entry	-0.15	0.045	-0.24 , -0.06	-0.05	0.032	-0.11 , 0.01	-0.22	0.125	-0.46 , 0.03	-0.22	0.070	-0.35 , -0.08
Use of statin in year prior to study entry	-0.13	0.036	-0.21 , -0.06	0.08	0.031	0.02 , 0.14	-0.2	0.097	-0.39 , -0.01	-0.18	0.073	-0.32 , -0.04
Use of NSAID in year prior to study entry	0.12	0.035	0.05 , 0.19	0.15	0.031	0.09 , 0.21	-0.1	0.101	-0.3 , 0.09	-0.04	0.077	-0.19 , 0.11
Use of Aspirin in year prior to study entry	-0.09	0.043	-0.18 , -0.01	0.06	0.035	-0.01 , 0.13	-0.16	0.116	-0.39 , 0.07	-0.02	0.078	-0.18 , 0.13

HbA1c at study entry												
<6%												
6% - 6.5%	-0.09	0.048	-0.19 , 0	0.48	0.042	0.4 , 0.56	0.01	0.141	-0.27 , 0.29	0.62	0.130	0.37 , 0.88
6.5%-7%	-0.18	0.051	-0.28 , -0.07	0.87	0.042	0.79 , 0.95	0.03	0.147	-0.26 , 0.31	1.07	0.126	0.82 , 1.32
7% - 8%	-0.24	0.052	-0.35 , -0.14	1.4	0.042	1.32 , 1.48	-0.11	0.153	-0.41 , 0.19	1.68	0.119	1.44 , 1.91
8%-10%	-0.24	0.063	-0.36 , -0.11	1.96	0.050	1.86 , 2.06	-0.18	0.171	-0.51 , 0.16	2.53	0.121	2.29 , 2.77
>10%	-0.26	0.081	-0.42 , -0.1	2.3	0.067	2.16 , 2.43	-0.22	0.209	-0.63 , 0.19	3.16	0.130	2.91 , 3.42
BMI at study entry												
<25												
25-29	0.12	0.070	-0.02 , 0.26	0.46	0.051	0.36 , 0.56	-0.35	0.130	-0.6 , -0.09	-0.59	0.074	-0.73 , -0.44
30-34	0.03	0.084	-0.13 , 0.2	0.56	0.052	0.46 , 0.66	-0.91	0.188	-1.28 , -0.54	-1.48	0.096	-1.66 , -1.29
35+	-0.14	0.104	-0.35 , 0.06	0.63	0.056	0.53 , 0.74	-1.67	0.307	-2.27 , -1.07	-1.91	0.127	-2.16 , -1.66
SBP at study entry												
100-129												
130-139	0.01	0.041	-0.07 , 0.09	0.06	0.041	-0.02 , 0.14	-0.1	0.098	-0.29 , 0.1	-0.11	0.092	-0.29 , 0.07
140-149	-0.03	0.043	-0.11 , 0.05	0.04	0.040	-0.04 , 0.12	0.08	0.099	-0.11 , 0.28	0.09	0.089	-0.09 , 0.26
>=150	-0.08	0.041	-0.16 , 0	0.02	0.039	-0.06 , 0.09	0.03	0.098	-0.17 , 0.22	0.09	0.085	-0.08 , 0.26
History of CVD at study entry	0.09	0.112	-0.13 , 0.3	0.01	0.048	-0.08 , 0.11	0.35	0.254	-0.14 , 0.85	0.41	0.105	0.21 , 0.62
History of CKD at study entry	-0.6	0.304	-1.19 , 0	-1.1	0.296	-1.68 , -0.52	0.58	0.299	0 , 1.17	1.31	0.270	0.78 , 1.83
History of Cancer at study entry	0.17	0.224	-0.27 , 0.61	0	0.228	-0.45 , 0.45	0.07	0.370	-0.66 , 0.79	0.42	0.355	-0.27 , 1.12
History of stroke at study entry	-0.2	0.174	-0.54 , 0.14	-0.01	0.083	-0.18 , 0.15	-0.65	0.361	-1.36 , 0.06	-0.39	0.203	-0.79 , 0.01
History of MI at study entry	0.17	0.206	-0.23 , 0.58	0.08	0.084	-0.09 , 0.24	-0.88	0.320	-1.51 , -0.25	0.32	0.169	-0.01 , 0.65
MI in three months before study entry	-0.26	0.266	-0.78 , 0.26	-0.43	0.225	-0.87 , 0.01	-0.01	0.416	-0.83 , 0.81	-0.22	0.434	-1.07 , 0.63
Stroke in previous three months	0.56	0.276	0.02 , 1.1				0.36	0.503	-0.63 , 1.34			
Any CVD event in three months before study entry	-0.01	0.084	-0.17 , 0.16	0.06	0.081	-0.1 , 0.22	-0.14	0.198	-0.53 , 0.25	-0.1	0.187	-0.47 , 0.27
TIME UPDATED												
Use of anti HT in previous year	0.24	0.049	0.15 , 0.34				0.09	0.131	-0.16 , 0.35			
Use of statin in previous year	0.4	0.037	0.32 , 0.47				0.04	0.094	-0.14 , 0.23			
Use of NSAID in previous year	-0.01	0.036	-0.08 , 0.06				0.08	0.099	-0.11 , 0.28			
Use of Aspirin in previous year	0.17	0.039	0.09 , 0.24				0.13	0.107	-0.08 , 0.34			
History of CVD	-0.17	0.107	-0.38 , 0.04				-0.08	0.249	-0.56 , 0.41			
History of CKD	-0.18	0.063	-0.3 , -0.05				0.8	0.136	0.53 , 1.06			
History of Cancer	-0.12	0.099	-0.31 , 0.08				0.44	0.194	0.06 , 0.82			
History of stroke	0.13	0.155	-0.17 , 0.44				0.33	0.325	-0.3 , 0.97			
History of MI	-0.17	0.190	-0.55 , 0.2				1.19	0.283	0.63 , 1.74			
Stroke in previous three months	0.56	0.276	0.02 , 1.1				0.36	0.503	-0.63 , 1.34			
MI in three months before study entry	-0.26	0.266	-0.78 , 0.26				-0.01	0.416	-0.83 , 0.81			
Any CVD event in previous three months	0.02	0.142	-0.26 , 0.3				0.25	0.264	-0.27 , 0.77			
HbA1c in previous interval												
<6%												
6% - 6.5%	0.75	0.185	0.39 , 1.12				0.87	0.403	0.08 , 1.66			

6.5%-7%	2.06	0.187	1.7 , 2.43				1.85	0.383	1.1 , 2.6			
7% - 8%	4.55	0.176	4.2 , 4.89				4.36	0.364	3.64 , 5.07			
8%-10%	6.89	0.181	6.53 , 7.24				6.78	0.360	6.08 , 7.49			
>10%	7.78	0.221	7.34 , 8.21				7.21	0.471	6.29 , 8.14			
HbA1c in previous interval (-2)												
<6%												
26% - 6.5%	0.12	0.162	-0.2 , 0.43				-0.28	0.349	-0.96 , 0.4			
26.5%-7%	-0.13	0.164	-0.45 , 0.19				-0.68	0.334	-1.33 , -0.02			
27% - 8%	-1.07	0.156	-1.38 , -0.77				-1.48	0.319	-2.11 , -0.86			
28%-10%	-2.36	0.165	-2.68 , -2.04				-2.47	0.324	-3.11 , -1.84			
2>10%	-2.7	0.212	-3.12 , -2.29				-2.13	0.454	-3.02 , -1.24			
Bmi in previous interval												
<25												
25-29	0.22	0.069	0.09 , 0.36				-0.37	0.129	-0.62 , -0.11			
30-34	0.47	0.082	0.31 , 0.63				-0.72	0.190	-1.09 , -0.35			
35+	0.72	0.103	0.52 , 0.92				-0.3	0.300	-0.89 , 0.29			
SBP in previous interval												
100-129												
130-139	0.03	0.038	-0.04 , 0.1				-0.11	0.091	-0.29 , 0.07			
140-149	0.03	0.039	-0.04 , 0.11				0.02	0.093	-0.16 , 0.2			
>=150	0.06	0.042	-0.02 , 0.15				0.12	0.097	-0.07 , 0.31			

Table 22.2 2005 onwards

	Metformin						Sulfonylurea					
	DENOMINATOR MODEL			**NUMERATOR MODEL**			**DENOMINATOR MODEL**			**NUMERATOR MODEL**		
	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI
Time (months) since study entry												
Time since study entry (months) spl 1	-0.03	0.001	-0.03 , -0.03	-0.04	0.001	-0.04 , -0.03	-0.03	0.004	-0.04 , -0.02	0.00	10.120	-0.03 , 0.00
Time since study entry (months) spl 2	0.03	0.001	0.02 , 0.03	0.03	0.001	0.03 , 0.04	0.02	0.006	0.01 , 0.03	0.04	0.005	0.03 , 0.05
Time (months) between diagnosis and study entry	-0.02	0.002	-0.02 , -0.02	-0.01	0.001	-0.01 , -0.01	-0.01	0.006	-0.03 , 0	-0.01	0.004	-0.01 , 0
Age at diagnosis (years)												
32-44												
45-49	-0.12	0.035	-0.19 , -0.05	-0.13	0.035	-0.2 , -0.06	-0.01	0.138	-0.28 , 0.26	-0.07	0.137	-0.34 , 0.2
60-74	-0.19	0.036	-0.26 , -0.12	-0.35	0.036	-0.42 , -0.28	0.04	0.144	-0.24 , 0.32	-0.13	0.144	-0.41 , 0.15
75-89	-0.49	0.044	-0.57 , -0.4	-0.7	0.044	-0.79 , -0.61	0.11	0.162	-0.21 , 0.42	0.02	0.159	-0.29 , 0.33
Gender	0.12	0.019	0.09 , 0.16	0.13	0.019	0.09 , 0.16	-0.1	0.073	-0.25 , 0.04	-0.1	0.072	-0.24 , 0.04
Smoking Status												
Non												
Current	0.08	0.027	0.02 , 0.13	0.13	0.026	0.08 , 0.18	0.2	0.093	0.01 , 0.38	0.29	0.091	0.11 , 0.47
Ex	0.07	0.020	0.03 , 0.11	0.08	0.021	0.04 , 0.12	0.08	0.077	-0.07 , 0.23	0.11	0.077	-0.05 , 0.26

Alcohol consumption												
non_drinker												
ex-drinker	-0.04	0.041	-0.13 , 0.04	-0.04	0.041	-0.12 , 0.04	0.06	0.136	-0.21 , 0.33	0.03	0.136	-0.24 , 0.29
current drinker unknown	-0.15	0.091	-0.33 , 0.03	-0.13	0.083	-0.29 , 0.03	-0.42	0.322	-1.05 , 0.21	-0.34	0.313	-0.95 , 0.28
rare drinker <2u/d	-0.06	0.032	-0.12 , 0	-0.05	0.032	-0.11 , 0.02	-0.13	0.113	-0.35 , 0.09	-0.14	0.111	-0.36 , 0.08
moderate drinker 3-6u/d	-0.08	0.030	-0.14 , -0.02	-0.09	0.031	-0.15 , -0.03	-0.2	0.107	-0.41 , 0.01	-0.24	0.107	-0.45 , -0.03
excessive drinker >6u/d	-0.12	0.040	-0.19 , -0.04	-0.14	0.041	-0.22 , -0.06	-0.04	0.145	-0.32 , 0.25	-0.13	0.145	-0.41 , 0.16
Use of anti HT in year prior to study entry	-0.15	0.039	-0.22 , -0.07	-0.03	0.021	-0.07 , 0.01	0.07	0.154	-0.23 , 0.37	0.22	0.080	0.06 , 0.37
Use of statin in year prior to study entry	-0.16	0.028	-0.21 , -0.1	0.25	0.021	0.21 , 0.29	-0.27	0.107	-0.48 , -0.06	-0.02	0.076	-0.17 , 0.13
Use of NSAID in year prior to study entry	0.03	0.027	-0.02 , 0.08	0.12	0.022	0.08 , 0.16	0.01	0.112	-0.21 , 0.23	0.22	0.083	0.06 , 0.38
Use of Aspirin in year prior to study entry	-0.11	0.036	-0.18 , -0.04	0.07	0.023	0.03 , 0.12	0.02	0.132	-0.24 , 0.28	0.25	0.085	0.09 , 0.42
HbA1c at study entry												
<6%												
6% - 6.5% base	-0.11	0.050	-0.21 , -0.02	0.6	0.040	0.52 , 0.68	-0.22	0.208	-0.62 , 0.19	0.47	0.176	0.13 , 0.82
6.5%-7% base	-0.23	0.051	-0.33 , -0.13	1.08	0.038	1.01 , 1.16	-0.34	0.212	-0.76 , 0.07	0.96	0.166	0.64 , 1.29
7% - 8% base	-0.32	0.053	-0.42 , -0.21	1.87	0.038	1.79 , 1.94	-0.33	0.219	-0.76 , 0.1	1.98	0.160	1.66 , 2.29
8%-10% base	-0.39	0.060	-0.51 , -0.27	2.58	0.044	2.5 , 2.67	-0.51	0.243	-0.99 , -0.04	2.93	0.170	2.6 , 3.26
>10% base	-0.37	0.075	-0.52 , -0.23	3.11	0.053	3.01 , 3.22	-0.21	0.261	-0.72 , 0.31	4.06	0.172	3.72 , 4.39
BMI at study entry												
<25												
25-29 base	-0.09	0.061	-0.21 , 0.03	0.26	0.037	0.19 , 0.34	-0.47	0.141	-0.75 , -0.2	-0.81	0.085	-0.98 , -0.65
30-34 base	-0.16	0.071	-0.3 , -0.02	0.35	0.038	0.27 , 0.42	-0.94	0.211	-1.35 , -0.52	-1.43	0.102	-1.63 , -1.24
35+ base	-0.28	0.082	-0.44 , -0.12	0.48	0.039	0.4 , 0.56	-1.04	0.314	-1.66 , -0.43	-1.63	0.118	-1.86 , -1.4
SBP at study entry												
100-129												
130-139base	-0.02	0.026	-0.07 , 0.03	0.02	0.024	-0.03 , 0.06	-0.09	0.095	-0.28 , 0.1	-0.18	0.087	-0.35 , -0.01
140-149base	-0.03	0.027	-0.09 , 0.02	0	0.025	-0.05 , 0.05	-0.05	0.105	-0.26 , 0.15	-0.1	0.090	-0.27 , 0.08
>=150base	-0.08	0.030	-0.14 , -0.02	0	0.027	-0.05 , 0.06	-0.11	0.114	-0.33 , 0.12	-0.08	0.097	-0.27 , 0.11
History of CVD at study entry	-0.03	0.101	-0.23 , 0.16	-0.05	0.037	-0.12 , 0.02	0.16	0.348	-0.53 , 0.84	0.13	0.124	-0.11 , 0.37
History of CKD at study entry	-0.17	0.059	-0.29 , -0.05	-0.35	0.038	-0.42 , -0.27	-0.14	0.158	-0.45 , 0.16	0.7	0.100	0.5 , 0.9
History of Cancer at study entry	-0.28	0.249	-0.77 , 0.21	-0.57	0.279	-1.11 , -0.02	-0.98	0.768	-2.48 , 0.53	-0.44	0.740	-1.89 , 1.01
History of stroke at study entry	-0.15	0.148	-0.44 , 0.14	-0.08	0.060	-0.2 , 0.03	-0.07	0.469	-0.99 , 0.85	-0.05	0.184	-0.41 , 0.31
History of MI at study entry	0.39	0.168	0.06 , 0.72	0.09	0.054	-0.02 , 0.19	-0.37	0.505	-1.36 , 0.62	0.08	0.179	-0.27 , 0.43
MI in three months before study entry	-0.13	0.204	-0.53 , 0.27	-0.24	0.208	-0.65 , 0.16	-0.15	0.649	-1.42 , 1.12	0.35	0.597	-0.82 , 1.52
Stroke in previous three months	-0.3	0.252	-0.79 , 0.2				-0.02	0.769	-1.53 , 1.49			
Any CVD event in three months before study entry	0.14	0.108	-0.07 , 0.35	0.13	0.115	-0.1 , 0.35	-0.06	0.420	-0.88 , 0.76	-0.08	0.397	-0.86 , 0.7
TIME UPDATED												
Use of anti HT in previous year	0.14	0.040	0.06 , 0.22				0.14	0.158	-0.17 , 0.45			
Use of statin in previous year	0.55	0.030	0.49 , 0.61				0.33	0.110	0.11 , 0.55			
Use of NSAID in previous year	0.1	0.027	0.05 , 0.15				0.29	0.112	0.07 , 0.51			
Use of Aspirin in previous year	0.22	0.034	0.15 , 0.28				0.29	0.129	0.03 , 0.54			

History of CVD	-0.09	0.097	-0.28 , 0.1				-0.16	0.340	-0.82 , 0.51			
History of CKD	-0.15	0.049	-0.24 , -0.05				1.01	0.142	0.73 , 1.28			
History of Cancer	-0.19	0.073	-0.33 , -0.04				0.64	0.187	0.27 , 1			
History of stroke	0.11	0.138	-0.16 , 0.38				0.1	0.439	-0.76 , 0.96			
History of MI	-0.36	0.161	-0.68 , -0.04				0.4	0.473	-0.53 , 1.33			
Stroke in previous three months	-0.3	0.252	-0.79 , 0.2				-0.02	0.769	-1.53 , 1.49			
MI in three months before study entry	-0.13	0.204	-0.53 , 0.27				-0.15	0.649	-1.42 , 1.12			
Any CVD event in previous three months	0.37	0.135	0.11 , 0.64				0.26	0.530	-0.78 , 1.3			
HbA1c in previous interval												
<6%												
6% - 6.5%	0.94	0.177	0.59 , 1.29				0.5	0.494	-0.47 , 1.47			
6.5%-7%	2.34	0.175	2 , 2.69				1.83	0.481	0.89 , 2.77			
7% - 8%	5.07	0.171	4.73 , 5.4				4.35	0.460	3.45 , 5.25			
8%-10%	7.53	0.174	7.19 , 7.87				7.29	0.463	6.38 , 8.19			
>10%	8.42	0.194	8.04 , 8.81				8.92	0.487	7.97 , 9.87			
HbA1c in previous interval (-2)												
<6%												
26% - 6.5%	-0.07	0.163	-0.4 , 0.25				0.04	0.437	-0.82 , 0.9			
26.5%-7%	-0.34	0.162	-0.66 , -0.03				-0.33	0.435	-1.19 , 0.52			
27% - 8%	-1.47	0.158	-1.78 , -1.16				-1.07	0.423	-1.9 , -0.24			
28%-10%	-2.73	0.163	-3.05 , -2.41				-2.3	0.440	-3.16 , -1.43			
2>10%	-3.07	0.188	-3.44 , -2.7				-3	0.469	-3.92 , -2.08			
Bmi in previous interval												
<25												
25-29	0.34	0.057	0.23 , 0.45				-0.47	0.141	-0.74 , -0.19			
30-34	0.47	0.067	0.34 , 0.61				-0.63	0.212	-1.05 , -0.22			
35+	0.74	0.079	0.58 , 0.89				-0.72	0.319	-1.35 , -0.09			
SBP in previous interval												
100-129												
130-139	0.02	0.024	-0.03 , 0.07				-0.22	0.092	-0.4 , -0.04			
140-149	0.05	0.027	-0.01 , 0.1				-0.11	0.103	-0.31 , 0.09			
>=150	0.12	0.032	0.05 , 0.18				0	0.117	-0.23 , 0.23			

Censoring models: Log odds ratios for being censored due to medication change (to something other than metformin or sulfonylureas), or transfer out. Estimated from a multinomial logistic regression model.

Table 22.3 2000-2005

	Medication change						Transfer out					
	DENOMINATOR MODEL			**NUMERATOR MODEL**			**DENOMINATOR MODEL**			**NUMERATOR MODEL**		
	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI
Medication in previous interval												
Metformin	1.41	0.092	1.23 , 1.59	2.08	0.091	1.9 , 2.26	-0.07	0.059	-0.18 , 0.05	-0.12	0.058	-0.23 , -0.01
Sulfonylurea	-0.02	0.085	-0.19 , 0.14	1.88	0.110	1.67 , 2.1	-0.02	0.085	-0.19 , 0.14	-0.05	0.085	-0.22 , 0.11
Time (months) since study entry												
Time since study entry (months) spl 1	0	0.004	-0.01 , 0.01	-0.01	0.004	-0.02 , 0	0	0.003	0 , 0.01	0	0.003	-0.01 , 0
Time since study entry (months) spl 2	0	0.003	-0.01 , 0	0.01	0.003	0 , 0.02	0	0.003	-0.01 , 0	0	0.003	0 , 0.01
Time (months) between diagnosis and study entry	-0.01	0.003	-0.01 , 0	0	0.003	-0.01 , 0.01	0	0.002	0 , 0.01	0.01	0.002	0 , 0.01
Age at diagnosis (years)												
32-44												
45-49	-0.26	0.065	-0.39 , -0.13	-0.46	0.062	-0.58 , -0.34	-0.31	0.082	-0.47 , -0.15	-0.39	0.081	-0.55 , -0.23
60-74	-0.52	0.074	-0.67 , -0.38	-0.93	0.069	-1.06 , -0.79	-0.35	0.087	-0.52 , -0.18	-0.44	0.083	-0.6 , -0.28
75-89	-1.06	0.121	-1.3 , -0.83	-1.45	0.115	-1.67 , -1.22	0.09	0.098	-0.1 , 0.28	0.07	0.094	-0.11 , 0.25
Gender	0.07	0.047	-0.02 , 0.16	0	0.045	-0.09 , 0.09	-0.14	0.048	-0.23 , -0.04	-0.13	0.048	-0.23 , -0.04
Smoking Status												
Non												
Current	0.14	0.055	0.03 , 0.25	0.24	0.054	0.13 , 0.34	0	0.063	-0.12 , 0.12	0.02	0.062	-0.1 , 0.15
Ex	0.11	0.050	0.02 , 0.21	0.09	0.049	0 , 0.19	0.04	0.05	-0.06 , 0.14	0.04	0.05	-0.06 , 0.13
Alcohol consumption												
non_drinker												
ex-drinker	0.05	0.125	-0.2 , 0.29	0.01	0.123	-0.23 , 0.25	-0.08	0.135	-0.35 , 0.18	-0.09	0.135	-0.36 , 0.17
current drinker unknown	0.03	0.137	-0.24 , 0.29	0.13	0.134	-0.14 , 0.39	0.42	0.121	0.18 , 0.66	0.45	0.121	0.21 , 0.69
rare drinker <2u/d	-0.05	0.076	-0.2 , 0.1	-0.06	0.075	-0.21 , 0.09	-0.04	0.079	-0.19 , 0.12	-0.04	0.078	-0.19 , 0.12
moderate drinker 3-6u/d	0.05	0.066	-0.08 , 0.19	0.02	0.066	-0.11 , 0.15	-0.02	0.069	-0.16 , 0.11	-0.04	0.069	-0.17 , 0.1
excessive drinker >6u/d	-0.03	0.097	-0.22 , 0.16	-0.2	0.096	-0.38 , -0.01	-0.01	0.1	-0.2 , 0.19	-0.03	0.1	-0.22 , 0.17
Use of anti HT in year prior to study entry	0.01	0.056	-0.1 , 0.12	-0.03	0.046	-0.12 , 0.06	0.21	0.071	0.07 , 0.35	-0.06	0.051	-0.16 , 0.04
Use of statin in year prior to study entry	-0.04	0.053	-0.14 , 0.06	-0.05	0.049	-0.15 , 0.05	0.08	0.056	-0.03 , 0.19	-0.04	0.051	-0.14 , 0.06
Use of NSAID in year prior to study entry	0.08	0.051	-0.02 , 0.18	0.11	0.049	0.02 , 0.21	0.06	0.057	-0.05 , 0.17	0.02	0.053	-0.08 , 0.12
Use of Aspirin in year prior to study entry	0.11	0.060	-0.01 , 0.23	0.14	0.055	0.04 , 0.25	0.05	0.064	-0.07 , 0.18	0.02	0.056	-0.09 , 0.13
HbA1c at study entry												
<6%												
6% - 6.5%	0.08	0.093	-0.1 , 0.26	0.22	0.092	0.05 , 0.4	0	0.076	-0.15 , 0.15	0	0.073	-0.15 , 0.14
6.5%-7%	-0.04	0.095	-0.23 , 0.15	0.12	0.094	-0.06 , 0.31	0.08	0.081	-0.08 , 0.24	0.06	0.075	-0.09 , 0.2

7% - 8%	0.01	0.088	-0.16 , 0.18	0.27	0.087	0.1 , 0.44	0.03	0.081	-0.13 , 0.18	0.02	0.075	-0.13 , 0.16
8%-10%	0.03	0.088	-0.14 , 0.2	0.41	0.088	0.24 , 0.59	0.04	0.087	-0.13 , 0.21	0.06	0.082	-0.1 , 0.22
>10%	0.1	0.092	-0.08 , 0.28	0.64	0.090	0.46 , 0.82	0.16	0.094	-0.03 , 0.34	0.16	0.09	-0.01 , 0.34
BMI at study entry												
<25												
25-29	-0.17	0.096	-0.35 , 0.02	-0.11	0.080	-0.27 , 0.04	0.17	0.085	0 , 0.33	-0.04	0.068	-0.17 , 0.09
30-34	-0.14	0.114	-0.37 , 0.08	0.11	0.082	-0.05 , 0.27	0.27	0.106	0.06 , 0.48	-0.03	0.073	-0.17 , 0.11
35+	-0.23	0.132	-0.49 , 0.03	0.18	0.086	0.01 , 0.34	0.18	0.136	-0.08 , 0.45	-0.07	0.082	-0.23 , 0.09
SBP at study entry												
100-129												
130-139	0.08	0.063	-0.05 , 0.2	0.06	0.061	-0.06 , 0.18	-0.01	0.064	-0.13 , 0.12	-0.05	0.064	-0.18 , 0.07
140-149	0.09	0.065	-0.04 , 0.21	0.06	0.061	-0.06 , 0.18	-0.14	0.067	-0.28 , -0.01	-0.21	0.065	-0.34 , -0.08
>=150	0.04	0.064	-0.09 , 0.16	0.01	0.060	-0.11 , 0.13	-0.1	0.065	-0.23 , 0.02	-0.17	0.062	-0.29 , -0.05
History of CVD at study entry	0.06	0.126	-0.19 , 0.31	0.08	0.084	-0.09 , 0.24	0.05	0.129	-0.2 , 0.3	0.07	0.081	-0.09 , 0.23
History of CKD at study entry	1.05	0.336	0.4 , 1.71	1.12	0.342	0.45 , 1.79	0.55	0.24	0.08 , 1.02	0.54	0.227	0.1 , 0.99
History of Cancer at study entry	-0.39	0.537	-1.44 , 0.67	-0.18	0.522	-1.2 , 0.85	-0.05	0.359	-0.75 , 0.66	-0.06	0.343	-0.74 , 0.61
History of stroke at study entry	-0.17	0.214	-0.59 , 0.25	-0.1	0.163	-0.42 , 0.22	-0.35	0.195	-0.73 , 0.03	0.08	0.135	-0.18 , 0.35
History of MI at study entry	0.37	0.239	-0.1 , 0.84	0.18	0.148	-0.12 , 0.47	-0.65	0.228	-1.09 , -0.2	-0.14	0.155	-0.44 , 0.16
MI in three months before study entry	0.46	0.364	-0.25 , 1.17	0.65	0.414	-0.16 , 1.46	-0.05	0.461	-0.95 , 0.85	-0.06	0.462	-0.96 , 0.85
Stroke in previous three months	-0.05	0.467	-0.96 , 0.87				0.49	0.389	-0.28 , 1.25			
Any CVD event in three months before study entry	-0.5	0.181	-0.86 , -0.15	-0.56	0.177	-0.9 , -0.21	-0.07	0.144	-0.35 , 0.22	-0.06	0.143	-0.34 , 0.22
TIME UPDATED												
Use of anti HT in previous year	0.02	0.063	-0.1 , 0.15				-0.46	0.075	-0.61 , -0.32			
Use of statin in previous year	0.14	0.055	0.03 , 0.25				-0.29	0.054	-0.4 , -0.18			
Use of NSAID in previous year	0.1	0.053	-0.01 , 0.2				-0.12	0.062	-0.24 , 0			
Use of Aspirin in previous year	0.06	0.048	-0.03 , 0.16				-0.07	0.055	-0.17 , 0.04			
History of CVD	-0.12	0.110	-0.33 , 0.1				0	0.118	-0.23 , 0.23			
History of CKD	0.27	0.068	0.13 , 0.4				0	0.071	-0.14 , 0.13			
History of Cancer	0.26	0.096	0.08 , 0.45				-0.02	0.106	-0.22 , 0.19			
History of stroke	0.16	0.156	-0.14 , 0.47				0.45	0.149	0.16 , 0.74			
History of MI	-0.14	0.194	-0.52 , 0.24				0.53	0.172	0.19 , 0.86			
Stroke in previous three months	-0.05	0.467	-0.96 , 0.87				0.49	0.389	-0.28 , 1.25			
MI in three months before study entry	0.46	0.364	-0.25 , 1.17				-0.05	0.461	-0.95 , 0.85			
Any CVD event in previous three months	0.52	0.267	0 , 1.05				0.25	0.261	-0.27 , 0.76			
HbA1c in previous interval												
<6%												
6% - 6.5%	0.3	0.390	-0.46 , 1.06				-0.1	0.133	-0.36 , 0.17			
6.5%-7%	0.63	0.383	-0.12 , 1.38				-0.44	0.182	-0.79 , -0.08			
7% - 8%	2.09	0.376	1.35 , 2.83				-0.56	0.204	-0.96 , -0.16			
8%-10%	3.45	0.372	2.72 , 4.18				-0.56	0.222	-1 , -0.13			
>10%	4.5	0.379	3.76 , 5.24				-0.34	0.29	-0.91 , 0.23			

HbA1c in previous interval (-2)												
<6%												
26% - 6.5%	-0.1	0.364	-0.81 , 0.61				0.12	0.136	-0.14 , 0.39			
26.5%-7%	0	0.355	-0.7 , 0.7				0.33	0.183	-0.03 , 0.69			
27% - 8%	-0.59	0.351	-1.28 , 0.1				0.47	0.206	0.06 , 0.87			
28%-10%	-0.95	0.349	-1.64 , -0.27				0.67	0.224	0.24 , 1.11			
2>10%	-1.39	0.358	-2.09 , -0.69				0.5	0.294	-0.07 , 1.08			
Bmi in previous interval												
<25												
25-29	-0.03	0.090	-0.2 , 0.15				-0.28	0.076	-0.43 , -0.13			
30-34	0.14	0.106	-0.07 , 0.34				-0.37	0.098	-0.56 , -0.18			
35+	0.27	0.125	0.03 , 0.52				-0.27	0.129	-0.52 , -0.01			
SBP in previous interval												
100-129												
130-139	-0.12	0.053	-0.22 , -0.01				-0.06	0.057	-0.17 , 0.05			
140-149	-0.14	0.057	-0.26 , -0.03				-0.05	0.061	-0.17 , 0.07			
>=150	-0.24	0.069	-0.37 , -0.11				-0.03	0.071	-0.17 , 0.1			

Table 22.4 2005 onwards

	Medication change						Transfer out					
	DENOMINATOR MODEL			**NUMERATOR MODEL**			**DENOMINATOR MODEL**			**NUMERATOR MODEL**		
	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI
Medication in previous interval												
Metformin	1.83	0.066	1.71 , 1.96	2.46	0.068	2.32 , 2.59	-0.09	0.042	-0.17 , -0.01	-0.12	0.042	-0.2 , -0.04
Sulfonylurea	1.27	0.098	1.07 , 1.46	2	0.103	1.8 , 2.21	-0.01	0.092	-0.19 , 0.17	-0.04	0.092	-0.22 , 0.14
Time (months) since study entry												
Time since study entry (months) spl 1	0	0.002	0 , 0.01	-0.01	0.002	-0.01 , 0	0	0.002	-0.01 , 0	0	0.002	-0.01 , 0
Time since study entry (months) spl 2	-0.01	0.002	-0.01 , 0	0	0.002	0 , 0.01	0	0.002	0 , 0.01	0	0.002	0 , 0.01
Time (months) between diagnosis and study entry	-0.02	0.005	-0.03 , -0.01	0	0.003	-0.01 , 0.01	0	0.002	0 , 0.01	0.01	0.002	0 , 0.01
Age at diagnosis (years)												
32-44												
45-49	-0.16	0.045	-0.25 , -0.07	-0.25	0.044	-0.34 , -0.17	-0.32	0.053	-0.43 , -0.22	-0.37	0.053	-0.47 , -0.26
60-74	-0.35	0.052	-0.45 , -0.25	-0.64	0.050	-0.74 , -0.54	-0.58	0.059	-0.7 , -0.47	-0.64	0.057	-0.75 , -0.53
75-89	-0.75	0.083	-0.91 , -0.59	-1.07	0.081	-1.23 , -0.91	-0.38	0.070	-0.52 , -0.24	-0.4	0.068	-0.54 , -0.27
Gender	0.08	0.033	0.02 , 0.15	0.03	0.033	-0.03 , 0.1	0.01	0.035	-0.06 , 0.08	0.01	0.035	-0.06 , 0.08
Smoking Status												
Non												
Current	0.07	0.042	-0.01 , 0.15	0.13	0.042	0.05 , 0.21	0.1	0.045	0.01 , 0.19	0.1	0.045	0.01 , 0.19
Ex	0.13	0.036	0.06 , 0.2	0.13	0.036	0.06 , 0.2	0.04	0.037	-0.03 , 0.11	0.03	0.037	-0.04 , 0.11
Alcohol consumption												

non_drinker												
ex-drinker	0.17	0.069	0.03 , 0.3	0.17	0.068	0.04 , 0.3	-0.12	0.072	-0.26 , 0.02	-0.12	0.072	-0.26 , 0.02
current drinker unknown	0.02	0.131	-0.23 , 0.28	0.07	0.130	-0.19 , 0.32	0.49	0.109	0.27 , 0.7	0.51	0.109	0.3 , 0.73
rare drinker <2u/d	0.05	0.054	-0.05 , 0.16	0.05	0.053	-0.05 , 0.16	-0.22	0.055	-0.33 , -0.11	-0.23	0.055	-0.33 , -0.12
moderate drinker 3-6u/d	0.05	0.052	-0.05 , 0.15	0.03	0.051	-0.07 , 0.13	-0.18	0.051	-0.28 , -0.08	-0.18	0.051	-0.28 , -0.08
excessive drinker >6u/d	0.04	0.071	-0.1 , 0.18	-0.04	0.070	-0.18 , 0.1	-0.17	0.070	-0.3 , -0.03	-0.17	0.070	-0.31 , -0.04
Use of anti HT in year prior to study entry	-0.04	0.049	-0.13 , 0.06	-0.03	0.035	-0.1 , 0.04	0.09	0.060	-0.03 , 0.2	-0.11	0.037	-0.19 , -0.04
Use of statin in year prior to study entry	0.02	0.039	-0.06 , 0.1	0.02	0.034	-0.05 , 0.08	0.12	0.044	0.04 , 0.21	-0.03	0.036	-0.1 , 0.04
Use of NSAID in year prior to study entry	0.08	0.040	0 , 0.15	0.08	0.037	0 , 0.15	0.02	0.044	-0.06 , 0.11	-0.04	0.041	-0.12 , 0.04
Use of Aspirin in year prior to study entry	0.01	0.051	-0.09 , 0.11	-0.03	0.041	-0.11 , 0.05	0.07	0.053	-0.03 , 0.17	0.03	0.041	-0.05 , 0.11
HbA1c at study entry												
<6%												
6% - 6.5%	-0.06	0.107	-0.27 , 0.15	0.17	0.104	-0.04 , 0.37	-0.04	0.069	-0.17 , 0.1	-0.08	0.064	-0.2 , 0.04
6.5%-7%	-0.11	0.103	-0.32 , 0.09	0.27	0.100	0.07 , 0.47	0.23	0.069	0.1 , 0.37	0.18	0.060	0.06 , 0.3
7% - 8%	-0.21	0.101	-0.41 , -0.01	0.35	0.099	0.16 , 0.55	0.22	0.073	0.08 , 0.36	0.19	0.063	0.07 , 0.32
8%-10%	-0.25	0.102	-0.45 , -0.05	0.51	0.100	0.32 , 0.71	0.2	0.078	0.04 , 0.35	0.21	0.071	0.07 , 0.35
>10%	-0.26	0.103	-0.46 , -0.06	0.73	0.101	0.53 , 0.92	0.08	0.086	-0.09 , 0.25	0.09	0.078	-0.07 , 0.24
BMI at study entry												
<25												
25-29	-0.12	0.091	-0.3 , 0.06	-0.22	0.065	-0.34 , -0.09	0.02	0.077	-0.13 , 0.17	-0.11	0.056	-0.22 , 0
30-34	-0.06	0.107	-0.27 , 0.15	-0.04	0.064	-0.17 , 0.08	0.07	0.093	-0.11 , 0.25	-0.14	0.058	-0.25 , -0.03
35+	0.04	0.118	-0.19 , 0.27	0.21	0.064	0.08 , 0.33	0.08	0.114	-0.15 , 0.3	-0.15	0.060	-0.27 , -0.04
SBP at study entry												
100-129												
130-139	-0.14	0.042	-0.22 , -0.06	-0.15	0.040	-0.23 , -0.07	-0.01	0.044	-0.1 , 0.08	-0.03	0.043	-0.11 , 0.05
140-149	-0.11	0.044	-0.19 , -0.02	-0.12	0.042	-0.21 , -0.04	0	0.047	-0.09 , 0.09	-0.03	0.045	-0.12 , 0.06
>=150	-0.11	0.049	-0.21 , -0.02	-0.15	0.046	-0.24 , -0.06	-0.02	0.052	-0.12 , 0.08	-0.03	0.048	-0.13 , 0.06
History of CVD at study entry	-0.04	0.145	-0.33 , 0.24	0.06	0.073	-0.08 , 0.2	0.2	0.139	-0.07 , 0.48	0.13	0.068	-0.01 , 0.26
History of CKD at study entry	0.15	0.098	-0.04 , 0.34	0.25	0.074	0.1 , 0.4	-0.02	0.091	-0.2 , 0.15	-0.03	0.068	-0.17 , 0.1
History of Cancer at study entry	0.87	0.423	0.04 , 1.7	1.32	0.379	0.58 , 2.07	-0.16	0.411	-0.97 , 0.64	-0.18	0.402	-0.96 , 0.61
History of stroke at study entry	-0.09	0.230	-0.54 , 0.36	-0.19	0.128	-0.44 , 0.06	-0.58	0.198	-0.97 , -0.19	-0.13	0.114	-0.36 , 0.09
History of MI at study entry	-0.2	0.221	-0.63 , 0.23	0.17	0.106	-0.04 , 0.38	-0.32	0.245	-0.8 , 0.16	-0.25	0.116	-0.48 , -0.02
MI in three months before study entry	0.51	0.341	-0.15 , 1.18	0.36	0.355	-0.34 , 1.05	-0.17	0.408	-0.97 , 0.63	-0.17	0.415	-0.98 , 0.64
Stroke in previous three months	-0.62	0.655	-1.9 , 0.66				1.13	0.413	0.32 , 1.94			
Any CVD event in three months before study entry	-0.21	0.227	-0.66 , 0.23	-0.23	0.222	-0.66 , 0.21	0.05	0.199	-0.34 , 0.44	0.06	0.197	-0.33 , 0.44
TIME UPDATED												
Use of anti HT in previous year	0.08	0.051	-0.02 , 0.18				-0.26	0.061	-0.38 , -0.14			
Use of statin in previous year	0.09	0.043	0.01 , 0.18				-0.26	0.045	-0.35 , -0.18			
Use of NSAID in previous year	0.04	0.044	-0.05 , 0.12				-0.15	0.050	-0.25 , -0.05			
Use of Aspirin in previous year	-0.08	0.047	-0.17 , 0.01				-0.06	0.050	-0.16 , 0.04			
History of CVD	0.05	0.133	-0.21 , 0.31				-0.08	0.132	-0.34 , 0.18			

History of CKD	0.18	0.072	0.04 , 0.32				-0.01	0.069	-0.14 , 0.13			
History of Cancer	0.12	0.098	-0.07 , 0.31				0	0.098	-0.2 , 0.19			
History of stroke	-0.05	0.196	-0.44 , 0.33				0.45	0.168	0.12 , 0.78			
History of MI	0.29	0.196	-0.09 , 0.68				0.07	0.223	-0.37 , 0.51			
Stroke in previous three months	-0.62	0.655	-1.9 , 0.66				1.13	0.413	0.32 , 1.94			
MI in three months before study entry	0.51	0.341	-0.15 , 1.18				-0.17	0.408	-0.97 , 0.63			
Any CVD event in previous three months	0.06	0.293	-0.52 , 0.63				-0.09	0.337	-0.75 , 0.57			
HbA1c in previous interval												
<6%												
6% - 6.5%	-0.03	0.347	-0.71 , 0.65				0.07	0.129	-0.19 , 0.32			
6.5%-7%	0.68	0.340	0.01 , 1.35				0.12	0.142	-0.16 , 0.4			
7% - 8%	2.42	0.339	1.76 , 3.09				-0.01	0.158	-0.32 , 0.31			
8%-10%	4.27	0.335	3.61 , 4.92				0.06	0.181	-0.3 , 0.41			
>10%	4.98	0.342	4.31 , 5.65				0.36	0.210	-0.05 , 0.77			
HbA1c in previous interval (-2)												
<6%												
26% - 6.5%	0.19	0.321	-0.44 , 0.81				-0.13	0.131	-0.39 , 0.12			
26.5%-7%	0.14	0.318	-0.48 , 0.76				-0.24	0.144	-0.52 , 0.04			
27% - 8%	-0.3	0.319	-0.92 , 0.33				-0.06	0.160	-0.37 , 0.25			
28%-10%	-1.25	0.316	-1.87 , -0.63				0.05	0.182	-0.31 , 0.41			
2>10%	-1.57	0.325	-2.21 , -0.94				-0.29	0.215	-0.71 , 0.13			
Bmi in previous interval												
<25												
25-29	-0.2	0.086	-0.37 , -0.03				-0.15	0.070	-0.29 , -0.02			
30-34	-0.13	0.102	-0.33 , 0.07				-0.24	0.088	-0.41 , -0.07			
35+	0.03	0.113	-0.2 , 0.25				-0.24	0.109	-0.45 , -0.03			
SBP in previous interval												
100-129												
130-139	-0.03	0.038	-0.11 , 0.04				-0.06	0.041	-0.14 , 0.02			
140-149	-0.08	0.044	-0.16 , 0.01				-0.07	0.047	-0.16 , 0.02			
>=150	-0.14	0.056	-0.25 , -0.03				0.1	0.057	-0.01 , 0.21			

APPENDIX 23 FULL RESULTS FOR CHAPTER 8 PRIMARY ANALYSIS, COVARIATE SPECIFICATION B: MI, STROKE AND ALL-CAUSE MORTALITY

Current medication only

MI

	Metformin			Sulfonylurea		
	Estimate	SE	95% CI	Estimate	SE	95% CI
Basic adjustment	1.10	0.08	0.95 , 1.26	1.29	0.16	1.00 , 1.65
Full baseline adjustment	0.99	0.08	0.84 , 1.17	1.05	0.15	0.79 , 1.38
Time updated adjustment	0.94	0.08	0.80 , 1.11	0.96	0.14	0.73 , 1.27
IPTW model	0.94	0.11	0.75 , 1.18	0.98	0.20	0.65 , 1.47
IPTW & PICW model	0.94	0.11	0.75 , 1.17	0.96	0.20	0.63 , 1.46

Table 23.1 HR for risk of MI with use of metformin or sulfonylurea (left, right respectively) compared to Diet only, estimated from 5 different models. 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 & IPCW.

Stroke

	Metformin			Sulfonylurea		
	Estimate	SE	95% CI	Estimate	SE	95% CI
Basic adjustment	1.02	0.06	0.91 , 1.14	1.00	0.11	0.80 , 1.25
Full baseline adjustment	1.09	0.07	0.95 , 1.24	1.07	0.13	0.84 , 1.36
Time updated adjustment	1.03	0.07	0.90 , 1.18	0.99	0.12	0.77 , 1.26
IPTW model	1.22	0.12	1.02 , 1.47	0.92	0.16	0.65 , 1.28
IPTW & IPCW model	1.24	0.12	1.03 , 1.49	0.95	0.16	0.68 , 1.33

Table 23.2 HR for risk of stroke with use of metformin or sulfonylurea (left, right respectively) compared to Diet only, estimated from 5 different models. 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 & IPCW.

All-cause mortality

	Metformin			Sulfonylurea		
	Estimate	SE	95% CI	Estimate	SE	95% CI
Basic adjustment	0.90	0.03	0.84 , 0.96	1.42	0.07	1.29 , 1.58
Full baseline adjustment	0.91	0.03	0.85 , 0.98	1.24	0.07	1.1 , 1.39
Time updated adjustment	0.98	0.04	0.91 , 1.06	1.30	0.08	1.15 , 1.47
IPTW model	0.94	0.05	0.83 , 1.05	1.08	0.10	0.90 , 1.29
IPTW & IPCW model	0.93	0.05	0.83 , 1.05	1.07	0.10	0.90 , 1.28

Table 23.3 HR for risk of all- cause mortality with use of metformin or sulfonylurea (left, right respectively) compared to Diet only, estimated from 5 different models. 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 & IPCW

HbA1c

	Metformin			Sulfonylurea		
	Absolute Difference in HbA1c (%)	SE	95% CI	Absolute Difference in HbA1c (%)	SE	95% CI
Basic adjustment	0.71	0.01	0.69 , 0.72	0.94	0.02	0.90 , 0.98
Full baseline adjustment	0.16	0.01	0.15 , 0.18	0.28	0.02	0.24 , 0.33
IPTW model	-0.27	0.01	-0.28 , 0.25	-0.15	0.01	-0.18 , 0.13
IPTW & IPCW model	-0.24	0.01	-0.26 ,0.23	-0.16	0.02	-0.19 , 0.13

Table 23.4 Absolute difference in HbA1c (%) with use of metformin (left) or sulfonylurea (right) compared to Diet only, estimated from 4 different models. 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 & IPCW. All models covariate specification B.

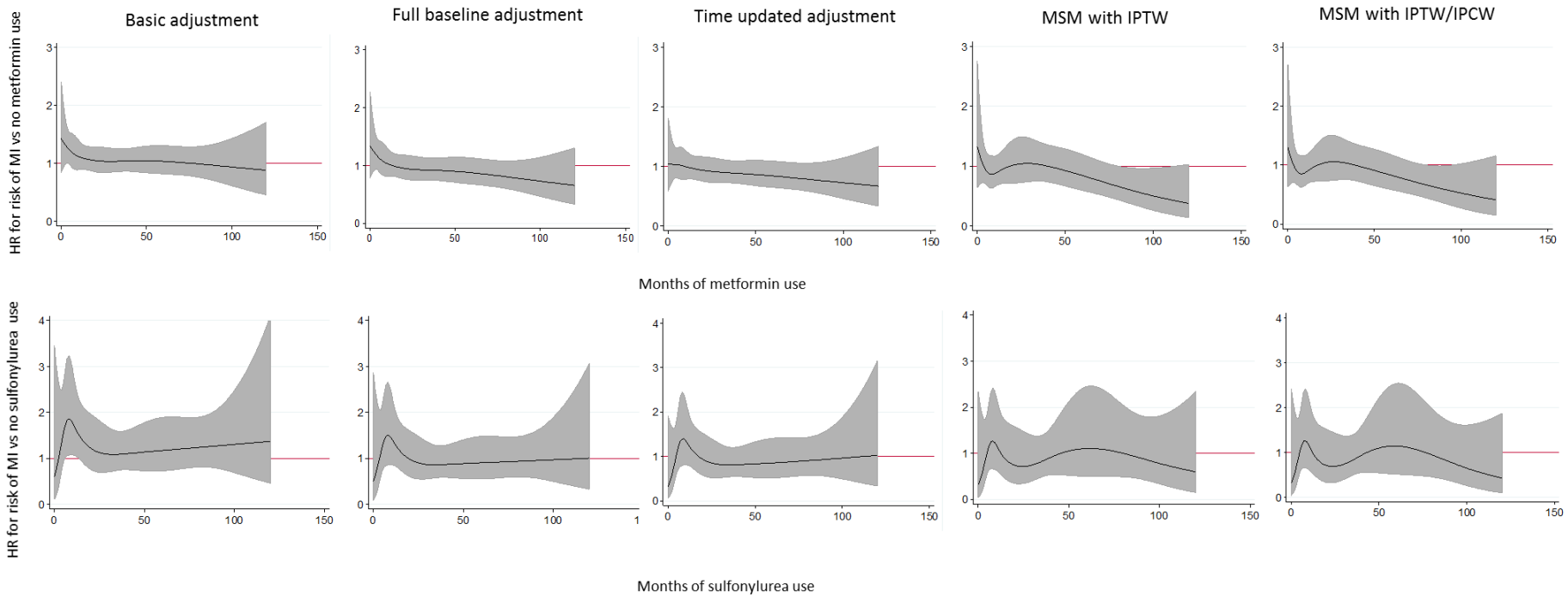
Cumulative use of medication

MI

	Metformin														
	Basic Adjustment			Baseline Adjustment			Time updated adjustment			IPTW			IPTW & IPCW		
	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI
< 3 month	1.33	0.24	0.94 , 1.88	1.24	0.22	0.87 , 1.76	1.02	0.19	0.71 , 1.48	1.20	0.32	0.71 , 2.01	1.17	0.31	0.7 , 1.96
3-6 months	1.18	0.23	0.81 , 1.71	1.09	0.21	0.74 , 1.6	0.97	0.19	0.66 , 1.43	0.84	0.19	0.53 , 1.31	0.82	0.19	0.52 , 1.29
6-12 months	1.13	0.17	0.85 , 1.51	1.04	0.16	0.77 , 1.4	1.00	0.15	0.75 , 1.35	0.87	0.15	0.62 , 1.21	0.86	0.15	0.61 , 1.21
1-2 years	1.13	0.13	0.9 , 1.43	1.02	0.13	0.8 , 1.31	1.01	0.13	0.79 , 1.29	1.06	0.20	0.72 , 1.55	1.08	0.21	0.74 , 1.57
2 - 5 years	1.06	0.11	0.88 , 1.29	0.93	0.10	0.75 , 1.15	0.90	0.10	0.72 , 1.11	1.00	0.16	0.73 , 1.37	0.99	0.16	0.72 , 1.35
> 5 years	0.88	0.13	0.67 , 1.17	0.71	0.11	0.52 , 0.97	0.70	0.11	0.51 , 0.95	0.59	0.14	0.37 , 0.93	0.60	0.14	0.38 , 0.95
	Sulfonylurea														
	Basic Adjustment			Baseline Adjustment			Time updated adjustment			IPTW			IPTW & IPCW		
	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI
< 3 month	0.92	0.53	0.3 , 2.87	0.76	0.44	0.24 , 2.36	0.55	0.32	0.17 , 1.73	0.65	0.38	0.2 , 2.04	0.65	0.38	0.21 , 2.07
3-6 months	1.31	0.66	0.49 , 3.53	1.07	0.55	0.39 , 2.9	0.89	0.46	0.32 , 2.43	0.58	0.39	0.16 , 2.14	0.57	0.38	0.15 , 2.09
6-12 months	1.98	0.61	1.08 , 3.6	1.59	0.50	0.86 , 2.93	1.49	0.46	0.81 , 2.72	1.18	0.43	0.58 , 2.41	1.18	0.43	0.58 , 2.41
1-2 years	1.36	0.38	0.78 , 2.37	1.09	0.31	0.62 , 1.92	1.04	0.30	0.59 , 1.83	1.12	0.51	0.47 , 2.72	1.09	0.49	0.46 , 2.61
2 - 5 years	1.03	0.23	0.67 , 1.59	0.81	0.19	0.51 , 1.26	0.75	0.17	0.48 , 1.18	0.67	0.20	0.37 , 1.21	0.68	0.20	0.38 , 1.21
> 5 years	1.27	0.29	0.81 , 2	0.96	0.24	0.6 , 1.56	0.93	0.23	0.57 , 1.52	1.13	0.46	0.51 , 2.52	1.08	0.46	0.46 , 2.51

Table 23.5 HR for risk of MI with cumulative use of metformin or sulfonylurea (Top, bottom respectively) compared to Diet only, estimated from 5 different models. 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 & IPCW.

Figure 23.1 HR for risk of MI with cumulative use of metformin or sulfonylurea (Top, bottom respectively) compared to Diet only. cumulative exposure modelled as cubic spline.

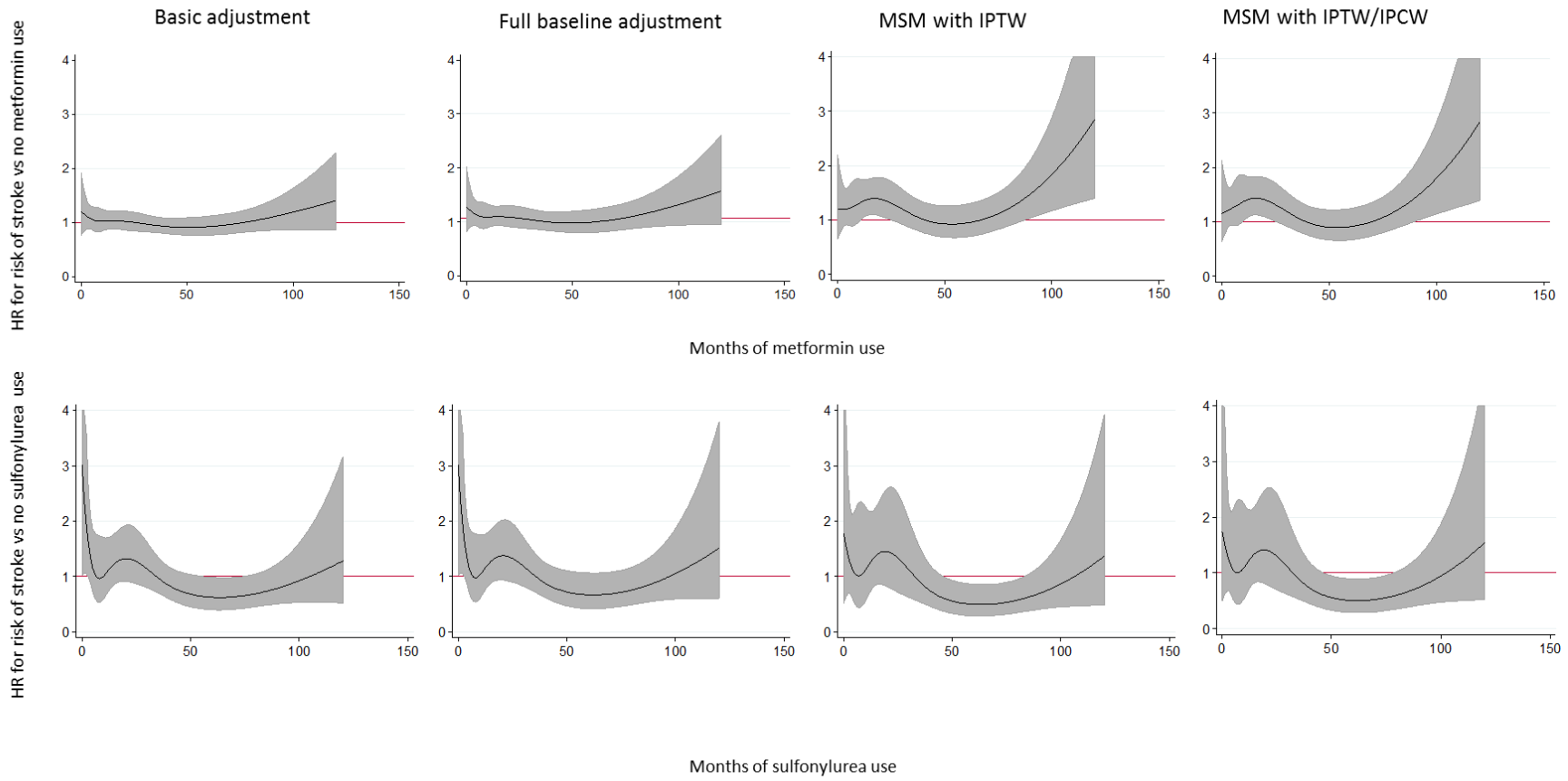


Stroke

	Metformin														
	Basic Adjust			Baseline Adjust			Time updated			IPTW			IPTW & IPCW		
	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI
< 3 month	1.32	0.19	0.99 , 1.76	1.39	0.21	1.03 , 1.86	1.08	0.18	0.79 , 1.48	1.39	0.26	0.97 , 2.01	1.40	0.26	0.97 , 2.01
3-6 months	0.77	0.15	0.52 , 1.13	0.81	0.16	0.55 , 1.19	0.77	0.15	0.52 , 1.14	0.85	0.24	0.50 , 1.47	0.90	0.26	0.51 , 1.6
6-12 months	1.07	0.14	0.84 , 1.37	1.14	0.15	0.88 , 1.46	1.12	0.15	0.87 , 1.45	1.35	0.27	0.91 , 2.00	1.43	0.29	0.96 , 2.14
1-2 years	1.02	0.10	0.83 , 1.24	1.08	0.11	0.88 , 1.33	1.06	0.11	0.86 , 1.3	1.32	0.21	0.96 , 1.81	1.37	0.22	1.00 , 1.87
2 - 5 years	0.97	0.08	0.82 , 1.14	1.03	0.10	0.86 , 1.24	0.98	0.09	0.82 , 1.17	1.09	0.15	0.83 , 1.44	1.09	0.15	0.83 , 1.43
> 5 years	0.99	0.11	0.8 , 1.24	1.06	0.13	0.83 , 1.36	1.03	0.13	0.81 , 1.32	1.17	0.19	0.85 , 1.61	1.15	0.19	0.83 , 1.58
	Sulfonylurea														
	Basic Adjust			Baseline Adjust			Time updated			IPTW			IPTW & IPCW		
	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI
< 3 month	1.80	0.61	0.93 , 3.49	1.78	0.60	0.92 , 3.46	1.23	0.44	0.61 , 2.46	1.23	0.47	0.58 , 2.58	1.22	0.47	0.57 , 2.59
3-6 months	0.84	0.42	0.32 , 2.25	0.84	0.42	0.31 , 2.26	0.80	0.41	0.3 , 2.17	1.27	0.91	0.31 , 5.14	1.21	0.87	0.29 , 4.98
6-12 months	1.25	0.38	0.69 , 2.28	1.26	0.39	0.69 , 2.3	1.27	0.39	0.7 , 2.32	1.14	0.39	0.58 , 2.23	1.15	0.39	0.59 , 2.25
1-2 years	1.30	0.30	0.83 , 2.03	1.33	0.31	0.84 , 2.1	1.31	0.31	0.83 , 2.07	1.10	0.34	0.61 , 2	1.07	0.32	0.59 , 1.94
2 - 5 years	0.91	0.17	0.63 , 1.31	0.96	0.19	0.65 , 1.4	0.89	0.17	0.61 , 1.31	0.94	0.27	0.53 , 1.66	0.97	0.28	0.55 , 1.69
> 5 years	0.71	0.17	0.45 , 1.13	0.80	0.19	0.5 , 1.28	0.75	0.18	0.47 , 1.21	0.54	0.15	0.32 , 0.92	0.59	0.17	0.34 , 1.03

Table 23.6 HR for risk of stroke with cumulative use of metformin or sulfonylurea (Top, bottom respectively) compared to Diet only, estimated from 5 different models. 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 & IPCW.

Figure 23.2 HR for risk of stroke with cumulative use of metformin or sulfonylurea (Top, bottom respectively) compared to Diet only. cumulative exposure modelled as cubic spline.

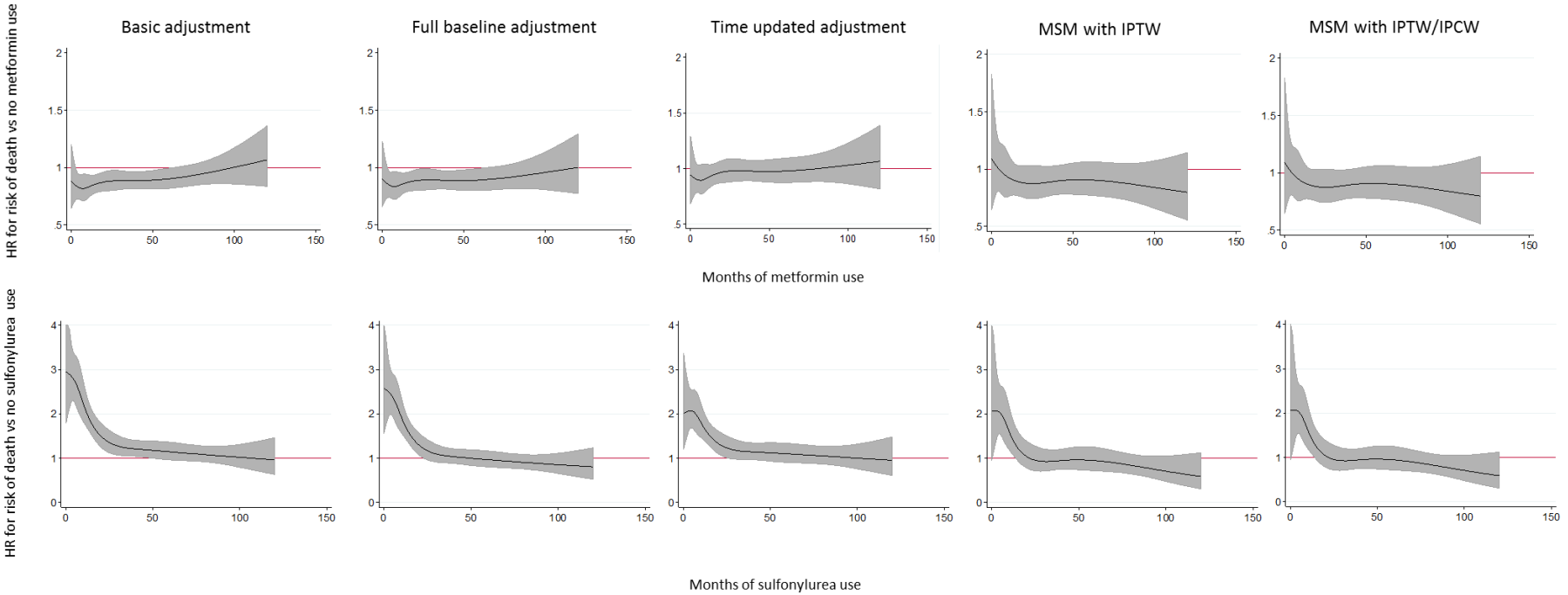


All-cause mortality

	Metformin														
	Basic Adjust			Baseline Adjust			Time updated			IPTW			IPTW & IPCW		
	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI
< 3 month	0.84	0.09	0.68 , 1.05	0.86	0.10	0.69 , 1.07	0.91	0.10	0.72 , 1.14	1.01	0.19	0.7 , 1.46	1.00	0.19	0.69 , 1.45
3-6 months	0.94	0.10	0.76 , 1.16	0.96	0.10	0.78 , 1.18	1.02	0.11	0.82 , 1.27	1.29	0.24	0.9 , 1.85	1.29	0.24	0.9 , 1.85
6-12 months	0.79	0.07	0.67 , 0.93	0.81	0.07	0.68 , 0.95	0.88	0.08	0.74 , 1.04	0.83	0.10	0.65 , 1.06	0.83	0.10	0.65 , 1.06
1-2 years	0.85	0.05	0.76 , 0.96	0.86	0.05	0.76 , 0.98	0.94	0.06	0.83 , 1.06	0.84	0.09	0.69 , 1.03	0.84	0.09	0.69 , 1.03
2 - 5 years	0.90	0.04	0.82 , 0.98	0.89	0.04	0.81 , 0.98	0.98	0.05	0.89 , 1.09	0.90	0.07	0.77 , 1.04	0.90	0.07	0.77 , 1.04
> 5 years	0.94	0.05	0.84 , 1.05	0.91	0.06	0.81 , 1.03	1.00	0.07	0.88 , 1.14	0.91	0.09	0.76 , 1.1	0.91	0.09	0.75 , 1.09
	Sulfonylurea														
	Basic Adjust			Baseline Adjust			Time updated			IPTW			IPTW & IPCW		
	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI	Est	SE	95% CI
< 3 month	2.86	0.46	2.08 , 3.94	2.50	0.41	1.81 , 3.45	2.03	0.34	1.46 , 2.83	2.13	0.50	1.34 , 3.38	2.10	0.49	1.33 , 3.33
3-6 months	2.70	0.45	1.94 , 3.75	2.35	0.39	1.69 , 3.26	2.02	0.35	1.44 , 2.82	1.69	0.38	1.08 , 2.62	1.61	0.36	1.04 , 2.51
6-12 months	2.44	0.31	1.91 , 3.13	2.12	0.27	1.65 , 2.73	1.97	0.26	1.53 , 2.55	1.79	0.30	1.3 , 2.48	1.80	0.30	1.3 , 2.49
1-2 years	1.53	0.18	1.22 , 1.92	1.32	0.16	1.05 , 1.67	1.33	0.16	1.05 , 1.68	1.16	0.23	0.78 , 1.73	1.15	0.23	0.78 , 1.7
2 - 5 years	1.23	0.10	1.05 , 1.45	1.05	0.09	0.89 , 1.25	1.16	0.11	0.97 , 1.39	0.96	0.13	0.73 , 1.25	0.95	0.13	0.73 , 1.23
> 5 years	1.04	0.10	0.86 , 1.25	0.87	0.09	0.72 , 1.06	1.02	0.11	0.82 , 1.25	0.80	0.12	0.59 , 1.07	0.81	0.12	0.6 , 1.08

Table 23.7HR for risk of all-cause mortality with cumulative use of metformin or sulfonylurea (Top, bottom respectively) compared to Diet only, estimated from5 different models. 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 & IPCW.

Figure 23.3 HR for risk of all-cause mortality with cumulative use of metformin or sulfonylurea (Top, bottom respectively) compared to Diet only. Cumulative exposure modelled as cubic spline

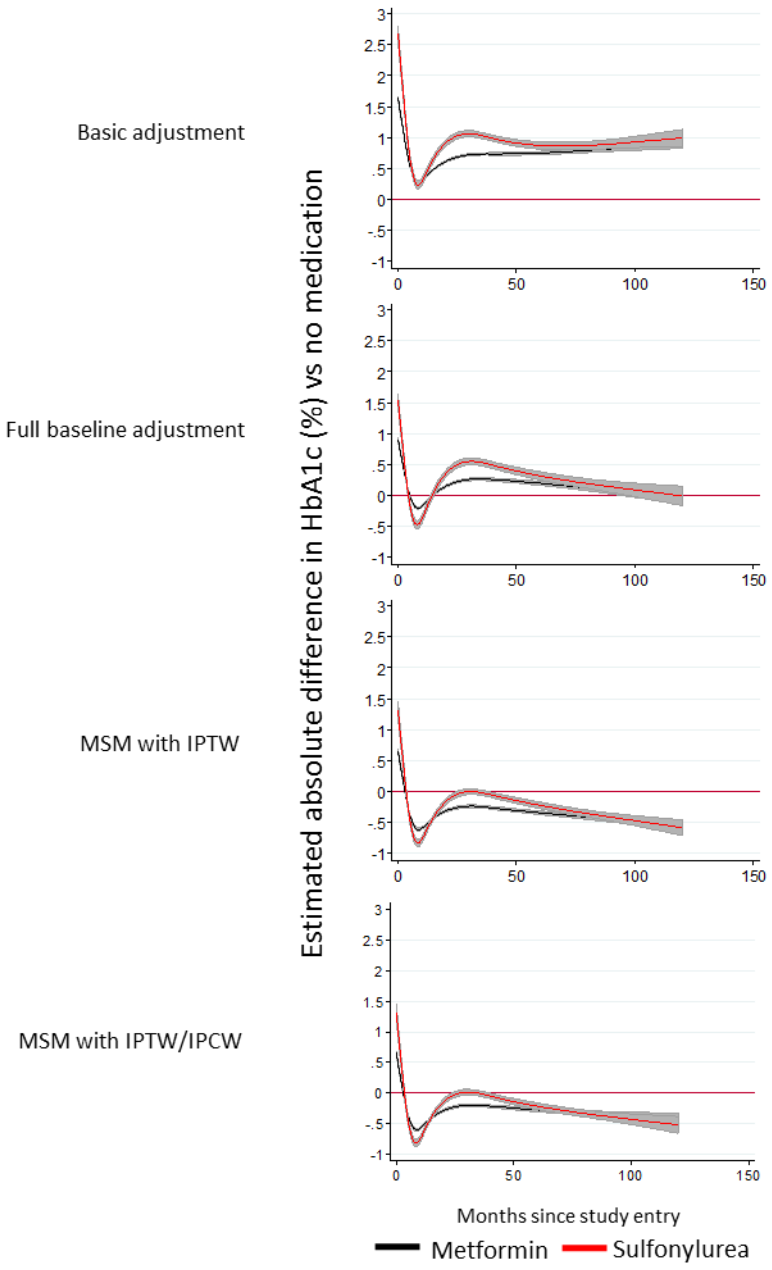


	Metformin											
	1 Basic Adjust			2 Baseline Adjust			4 IPTW			5 IPTW & IPCW		
	Est*	SE	95% CI	Est*	SE	95% CI	Est*	SE	95% CI	Est*	SE	95% CI
< 3 month	1.12	0.01	1.09 , 1.14	0.47	0.01	0.45 , 0.49	0.16	0.01	0.14 , 0.19	0.16	0.01	0.14 , 0.19
3-6 months	0.42	0.01	0.4 , 0.45	-0.12	0.01	-0.14 , -0.09	-0.49	0.02	-0.52 , -0.46	-0.49	0.02	-0.52 , -0.46
6-12 months	0.46	0.01	0.44 , 0.48	-0.04	0.01	-0.06 , -0.02	-0.45	0.01	-0.48 , -0.43	-0.45	0.01	-0.47 , -0.42
1-2 years	0.58	0.01	0.56 , 0.6	0.10	0.01	0.08 , 0.12	-0.35	0.01	-0.37 , -0.32	-0.32	0.01	-0.35 , -0.3
2 - 5 years	0.72	0.01	0.7 , 0.74	0.24	0.01	0.22 , 0.26	-0.27	0.01	-0.3 , -0.25	-0.23	0.01	-0.26 , -0.21
> 5 years	0.80	0.02	0.76 , 0.83	0.17	0.02	0.13 , 0.21	-0.37	0.02	-0.4 , -0.33	-0.29	0.02	-0.33 , -0.25
	Sulfonylurea											
	1 Basic Adjust			2 Baseline Adjust			4 IPTW			5 IPTW & IPCW		
	Est*	SE	95% CI	Est*	SE	95% CI	Est*	SE	95% CI	Est*	SE	95% CI
< 3 month	1.76	0.04	1.67 , 1.84	0.76	0.04	0.69 , 0.83	0.51	0.05	0.41 , 0.61	0.51	0.05	0.41 , 0.61
3-6 months	0.37	0.04	0.3 , 0.44	-0.46	0.04	-0.54 , -0.38	-0.75	0.05	-0.84 , -0.65	-0.74	0.05	-0.83 , -0.64
6-12 months	0.61	0.03	0.54 , 0.67	-0.10	0.04	-0.17 , -0.03	-0.51	0.04	-0.58 , -0.44	-0.50	0.04	-0.57 , -0.43
1-2 years	0.85	0.03	0.79 , 0.91	0.25	0.03	0.19 , 0.31	-0.22	0.03	-0.28 , -0.16	-0.20	0.03	-0.26 , -0.14
2 - 5 years	0.96	0.03	0.91 , 1.01	0.46	0.03	0.4 , 0.51	-0.09	0.02	-0.13 , -0.04	-0.08	0.02	-0.13 , -0.03
> 5 years	0.89	0.03	0.82 , 0.96	0.23	0.04	0.15 , 0.31	-0.33	0.03	-0.38 , -0.27	-0.31	0.03	-0.37 , -0.26

Table 23.8 Absolute difference in HbA1c (%) for cumulative use of metformin or sulfonylurea (Top, bottom respectively) relative to no use (diet only), estimated from 4 different models. 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 & IPCW. All models covariate specification B.

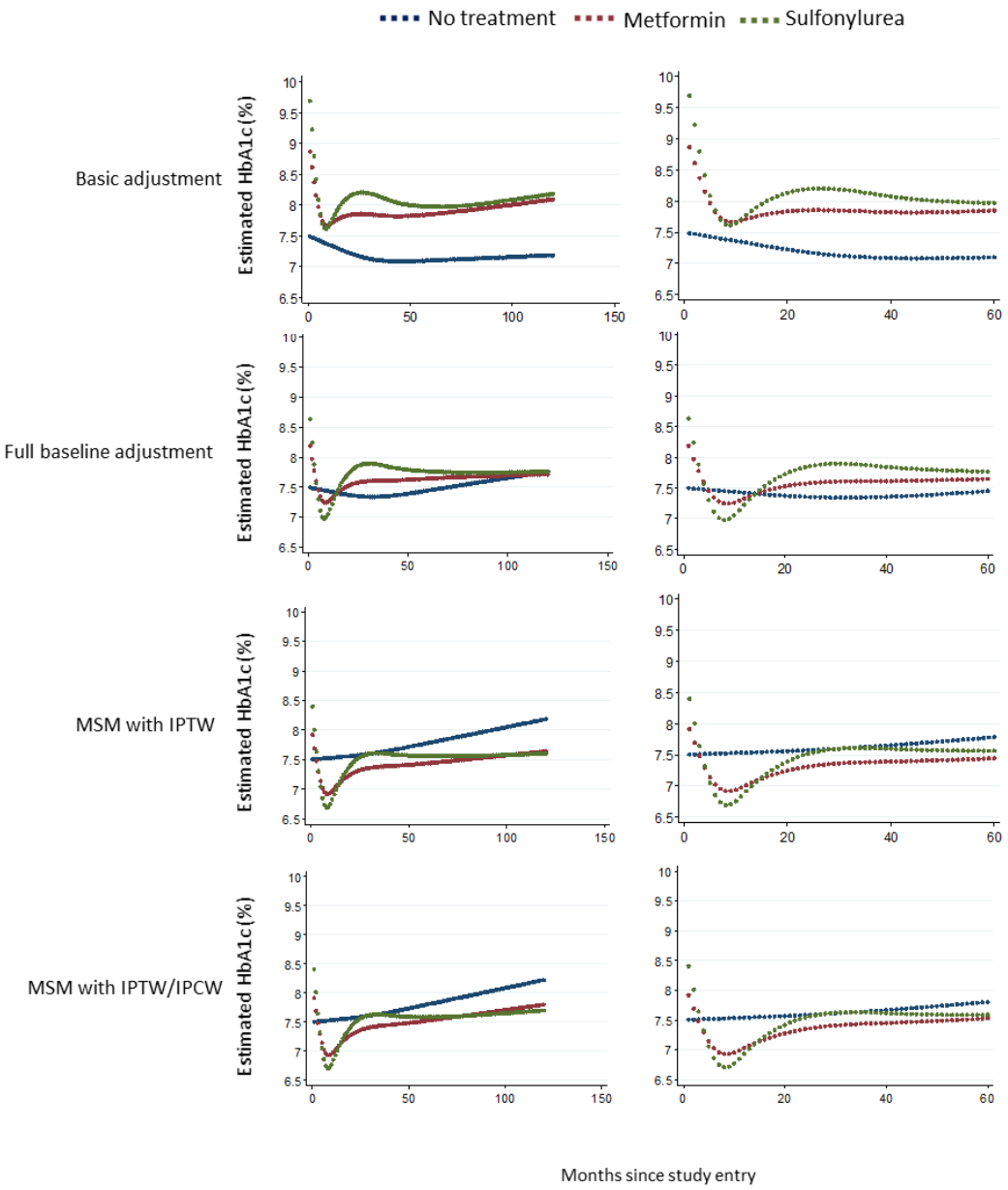
*Est: absolute difference in HbA1c

Figure 23.4 Absolute Difference in HbA1c (%) compared to no medication (diet only) with continued use of metformin (black) or sulfonylurea (red). All models covariate specification B (categorical)



95% CIs given by grey shading.

Figure 23.5 Estimated trajectory of HbA1c through time on the three treatment options. All models covariate specification B (categorical). Left – full follow up, Right – first 5 years only



APPENDIX 24 FULL RESULTS OF CHAPTER 8 SENSITIVITY ANALYSES

Figure 24.1 Comparison of HR estimates from primary analysis (solid lines) and sensitivity analysis in younger overweight population (dashed lines)

HR's displayed are from MSM with IPTW only using covariate specification A. HR's are presented for risk of MI (top), stroke (middle) and all-cause mortality (bottom)

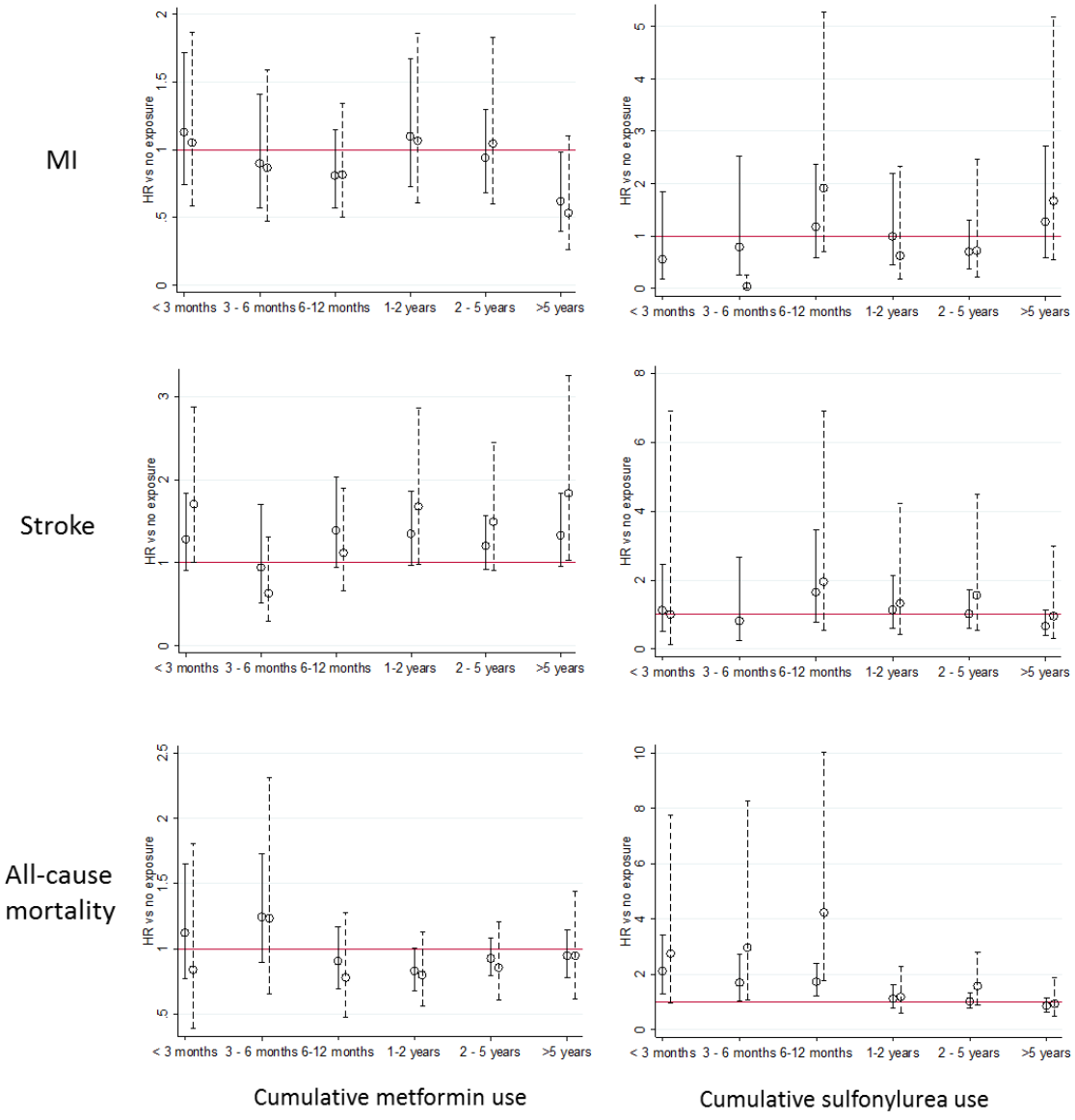
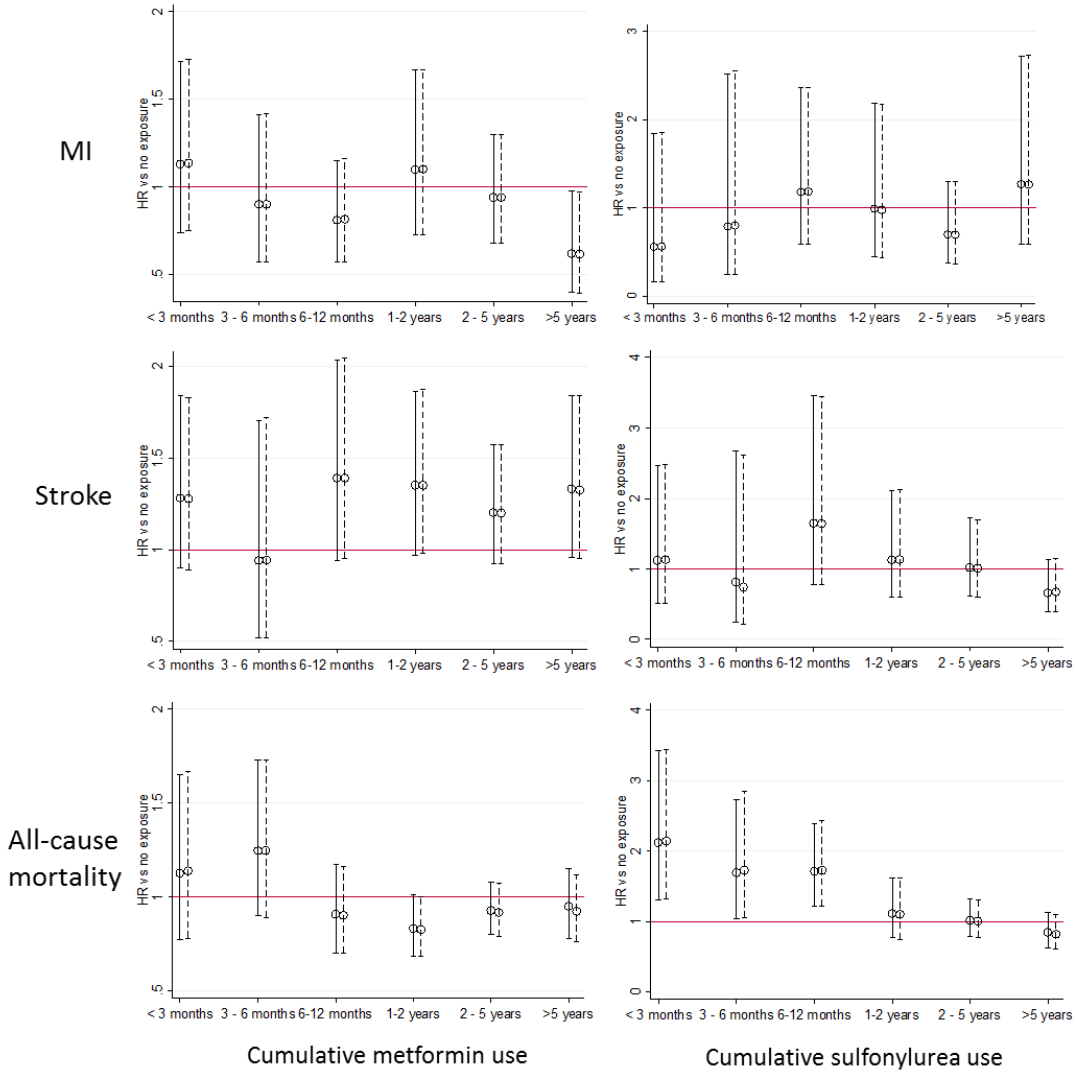


Figure 24.2 Comparison of HR estimates from primary analysis (solid lines) and sensitivity analysis where patients excluded if they have a history of cancer at study entry (dashed lines)



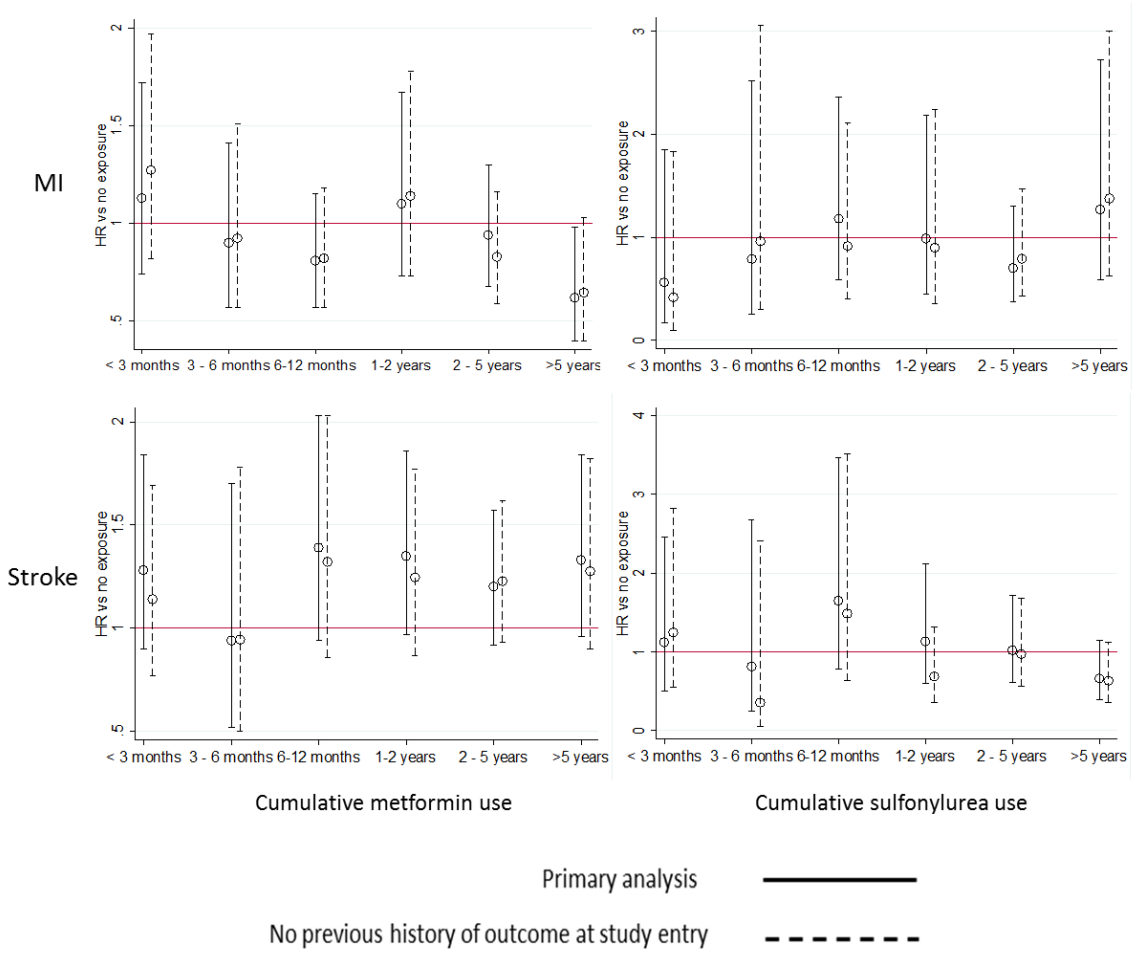
HR's displayed are from MSM with IPTW only using covariate specification A. HR's are presented for risk of MI (top), stroke (middle) and all-cause mortality (bottom)

		Primary analysis			BMI >=25 & Age < 65			No history of cancer at study entry		
	Length of exposure	**Est***	**SE**	**95% CI**	**Est***	**SE**	**95% CI**	**Est***	**SE**	**95% CI**
Metformin	**1- 3 month**	-0.1	0.01	-0.12 , -0.07	-0.08	0.02	-0.12 , -0.05	-0.10	0.01	-0.12 , -0.07
	3-6 months	-0.44	0.02	-0.47 , -0.41	-0.47	0.02	-0.51 , -0.43	-0.44	0.02	-0.47 , -0.41
	6-12 months	-0.22	0.01	-0.25 , -0.19	-0.24	0.02	-0.28 , -0.20	-0.22	0.01	-0.25 , -0.19
	12 - 24 months	-0.2	0.01	-0.23 , -0.18	-0.25	0.02	-0.28 , -0.21	-0.20	0.01	-0.23 , -0.18
	2 - 5 years	-0.33	0.01	-0.35 , -0.30	-0.46	0.02	-0.49 , -0.42	-0.33	0.01	-0.35 , -0.3
	> 5 years	-0.26	0.02	-0.30 , -0.22	-0.43	0.03	-0.49 , -0.37	-0.26	0.02	-0.30 , -0.22
	Length of exposure	**Est***	**SE**	**95% CI**	**Est***	**SE**	**95% CI**	**Est***	**SE**	**95% CI**
Sulfonylureas	**1-3 month**	-0.16	0.04	-0.25 , -0.07	-0.15	0.08	-0.31 , 0.01	-0.16	0.04	-0.25 , -0.07
	3-6 months	-0.73	0.06	-0.85 , -0.61	-0.66	0.12	-0.90 , -0.43	-0.73	0.06	-0.85 , -0.61
	6-12 months	-0.19	0.04	-0.27 , -0.11	-0.20	0.07	-0.34 , -0.06	-0.19	0.04	-0.27 , -0.11
	12 - 24 months	0.05	0.03	-0.02 , 0.11	0.04	0.06	-0.08 , 0.16	0.04	0.03	-0.02 , 0.11
	2 - 5 years	-0.18	0.02	-0.22 , -0.13	-0.32	0.04	-0.41 , -0.23	-0.18	0.02	-0.23 , -0.13
	> 5 years	-0.22	0.03	-0.28 , -0.16	-0.38	0.05	-0.48 , -0.28	-0.22	0.03	-0.28 , -0.17

Table 24.1 Absolute difference in HbA1c (%) with use of metformin or sulfonylurea (left, right respectively) compared to Diet only, estimated from a MSM with IPTW only, using covariate specification A. Results are from the primary analysis (left).; a population restricted to age<65 and BMI>=25 at study entry (middle); and a population restricted to have no history of cancer at study entry (right).

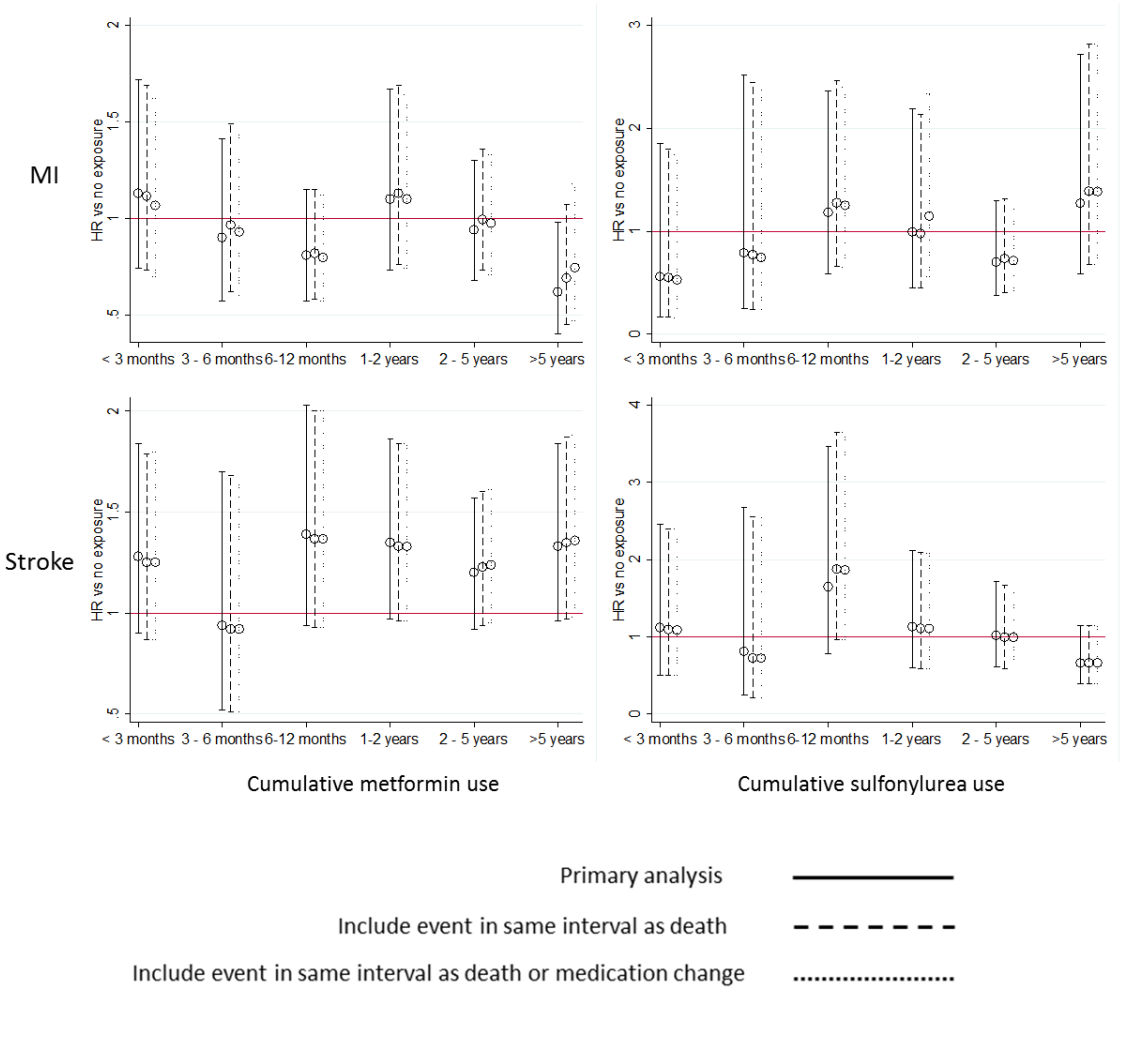
*Est: absolute difference in HbA1c (%)

Figure 24.3 Comparison of HR estimates from primary analysis (solid lines) and sensitivity analysis where population are restricted to have no history of the outcome at study entry (dashed lines)



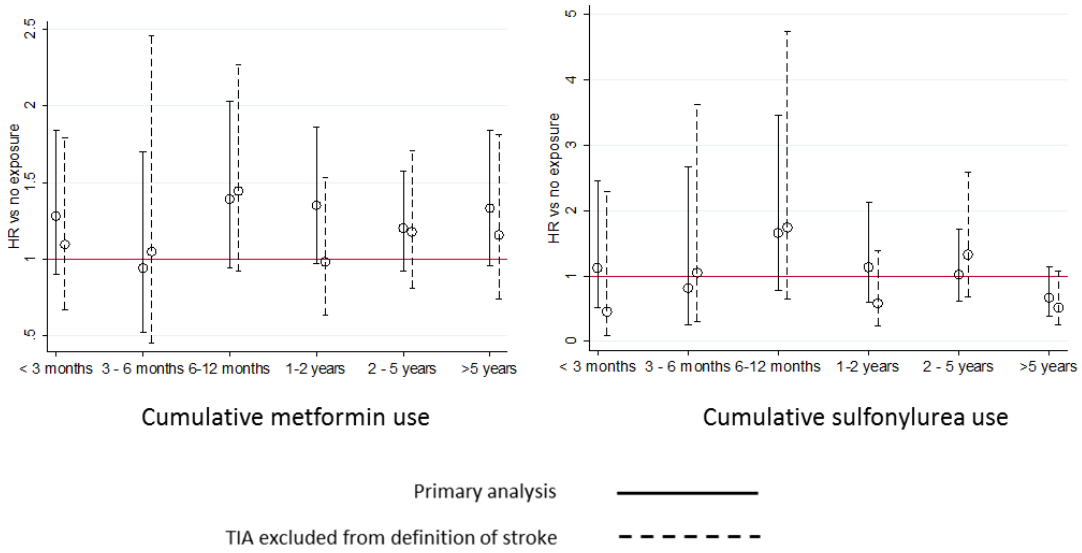
HR's displayed are from MSM with IPTW only using covariate specification A. HR's are presented for risk of MI (top), and stroke (bottom).

Figure 24.4 Comparison of HR estimates from primary analysis (solid lines) and sensitivity analysis where a subject is not censored if they have the outcome in the same interval as death (dashed lines) or in the same interval as death or medication change (dotted).



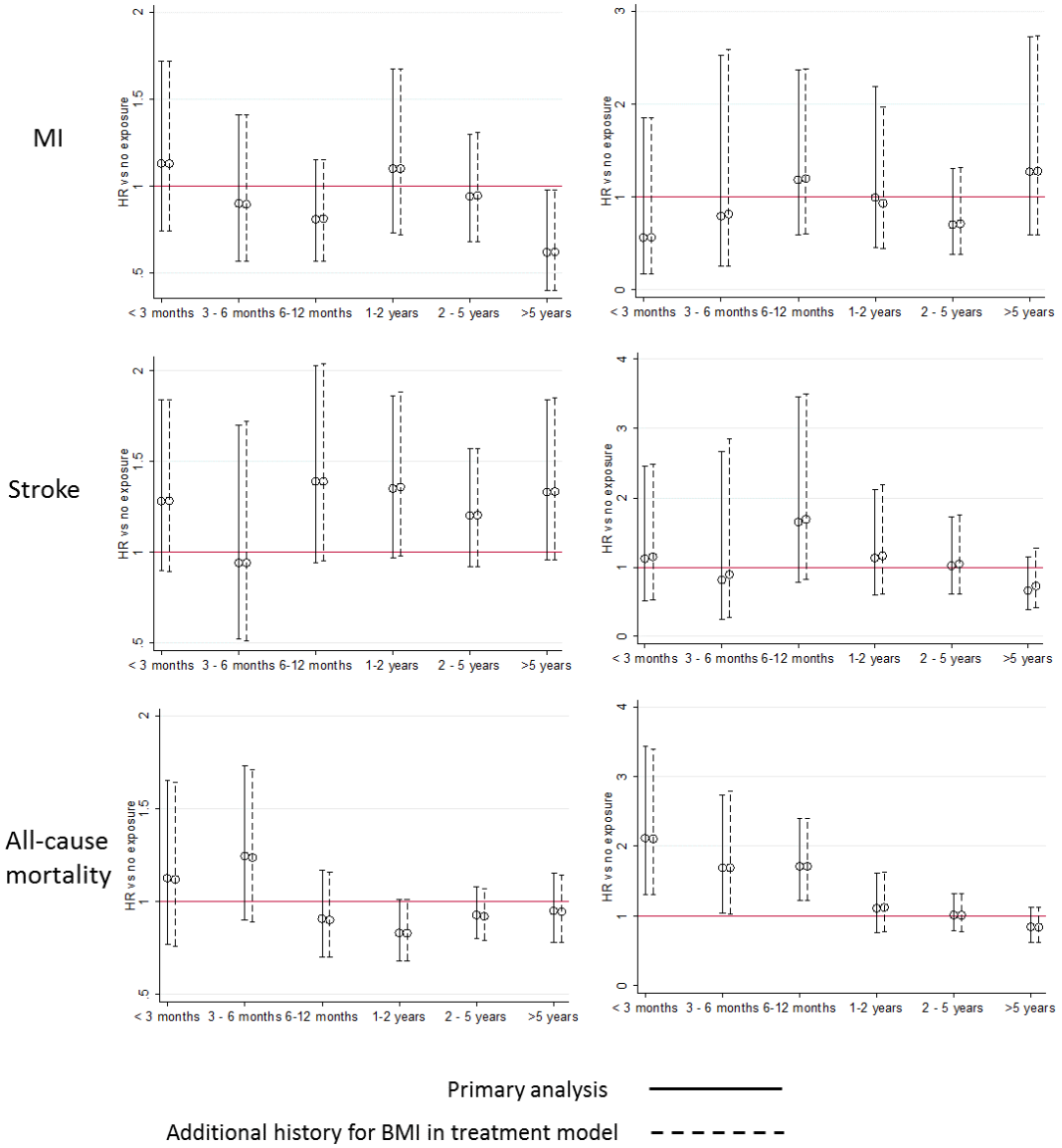
HR's displayed are from MSM with IPTW only using covariate specification A. HR's are presented for risk of MI (top), and stroke (bottom).

Figure 24.5 Comparison of HR estimates from primary analysis (solid lines) and sensitivity analysis where TIA is removed from the definition of stroke (dashed lines)



HR's displayed are from MSM with IPTW only using covariate specification A.

Figure 24.6 Comparison of HR estimates from primary analysis (solid lines) and sensitivity analysis where an additional term to better model short term change in BMI is included in the weighting model (dashed lines)



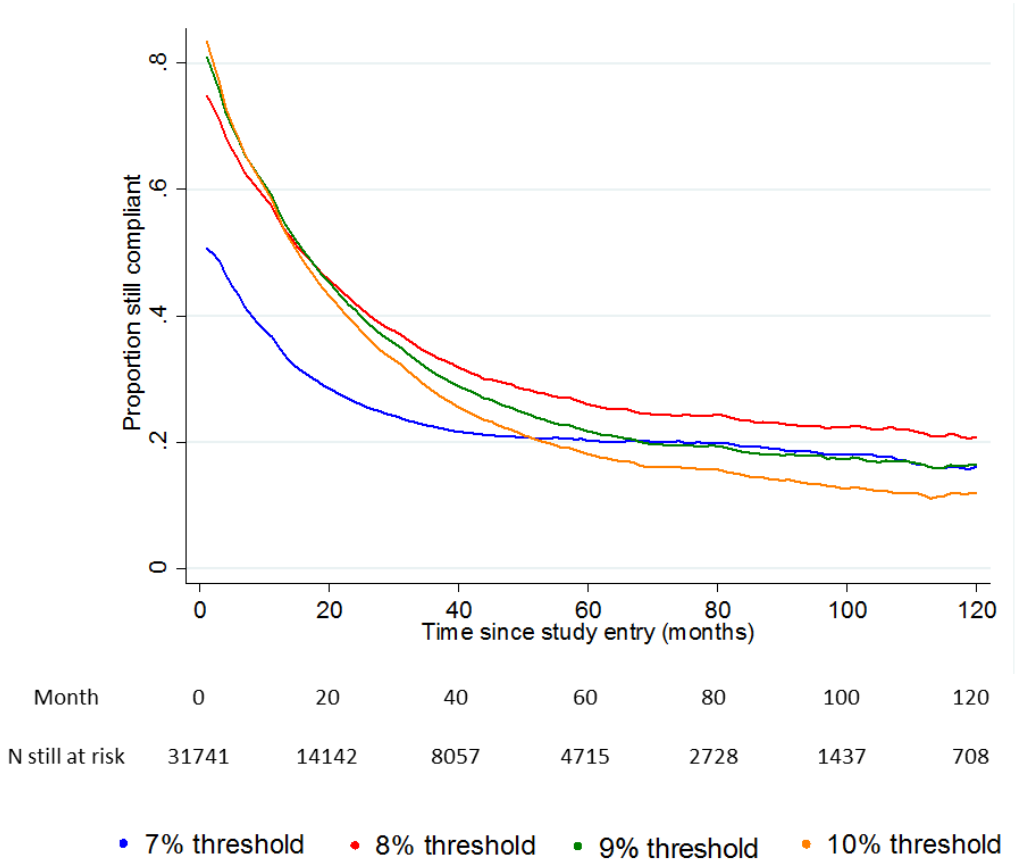
HR's displayed are from MSM with IPTW only using covariate specification A. HR's are presented for risk of MI (top), and stroke (middle), and all-cause mortality (bottom).

APPENDIX 25 EXTRA DESCRIPTIVE ANALYSES FOR TIME TO TARGET HBA1C OF 6.5%

Covariate	Mean	SD	Median, IQR
Age at diabetes diagnosis	61.5	12.1	62 (53 - 71)
Time between diagnosis and study entry (months)	2.8	9.4	0 (0 - 2)
A1c at study entry	7.9	1.6	7.3 (6.8 - 8.4)
BMI at study entry	32.1	6.1	31.2 (27.8 - 35.6)
SBP at study entry	137.8	16.3	138 (128 - 147)
Sex			
Male	18,015	57%	
Female	13,726	43%	
History of cancer *			
No	31,645	99.7%	
Yes	96	0.3%	
History of CVD			
No	27,112	85%	
Yes	4,629	15%	
CVD event in past 3 months			
No	31,166	98%	
Yes	575	2%	
History of MI			
No	30,566	96%	
Yes	1,175	4%	
MI in past 3 months			
No	31,639	99.7%	
Yes	102	0.3%	
History of Stroke			
No	30,659	97%	
Yes	1,082	3%	
Stroke in past 3 months			
No	31,617	99.6%	
Yes	124	0.39%	
History of CKD			
No	29,766	94%	
Yes	1,975	6%	
Use of statins in previous year			
No	15,641	49%	
Yes	16,100	51%	
Use of anti HTs in previous year			
No	12,432	39%	
Yes	19,309	61%	
Use of NSAIDS in previous year			
No	25,642	81%	
Yes	6,099	19%	
Use of ASPIRIN in previous year			
No	23,146	73%	
Yes	8,595	27%	
Smoking Status			
non	12,605	40%	
current	5,992	19%	
ex	13,144	41%	
Alcohol consumption			
non-drinker	3,830	12%	
ex-drinker	2,559	8%	
current drinker unknown	576	2%	
rare drinker <2u/d	7,960	25%	
moderate drinker 3-6u/d	14,133	45%	
excessive drinker >6u/d	2,683	8%	
Year of diabetes onset			
2000-2005	7,378	23.2%	
post 2005	24,363	76.8%	

Table 25.1 Cohort demographics at time of study entry for population relevant to outcome of reaching target HbA1c For clarity, the patients included in these analyses are on no medication at study entry; must have an HbA1c of greater than 6.5% to enter the study; and are lost from follow up at the point they reach their target HbA1c of 6.5% or less.

Figure 25.1 Proportion of patients under follow up who are still compliant to the treatment strategy "treat with metformin or sulfonylurea when HbA1c first rises above x%" at each month of follow up, for x = 7, 8, 9 and 10. Number still at risk (denominator) displayed below x-axis.



APPENDIX 26 ADDITIONAL TABLES OF RESULTS USING 3 MONTH GRACE PERIOD FOR ALL OUTCOMES

	Unweighted model baseline adjusted – hazard ratio for strategy vs 7% for reaching target HbA1c: three 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%	0.94 (0.93 , 0.95)	0.87 (0.85 , 0.90)	0.86 (0.83 , 0.89)	0.88 (0.83 , 0.95)	0.99 (0.91 , 1.10)	0.93 (0.85 , 1.01)
9%	0.93 (0.92 , 0.94)	0.85 (0.82 , 0.87)	0.82 (0.79 , 0.86)	0.84 (0.79 , 0.90)	0.86 (0.77 , 0.99)	0.88 (0.78 , 0.98)
10%	0.92 (0.91 , 0.94)	0.84 (0.82 , 0.87)	0.80 (0.77 , 0.85)	0.79 (0.73 , 0.87)	0.79 (0.68 , 0.94)	0.80 (0.70 , 0.90)
	IPW* Dynamic MSM – hazard ratio for strategy vs 7% for reaching target HbA1c: three 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%	0.97 (0.93 , 1.01)	0.85 (0.79 , 0.92)	0.80 (0.72 , 0.88)	0.76 (0.66 , 0.92)	1.02 (0.84 , 1.25)	0.93 (0.75 , 1.13)
9%	0.97 (0.92 , 1.00)	0.79 (0.72 , 0.85)	0.71 (0.64 , 0.8)	0.72 (0.60 , 0.88)	0.68 (0.54 , 0.87)	0.81 (0.63 , 1.05)
10%	0.98 (0.94 , 1.02)	0.80 (0.73 , 0.87)	0.68 (0.61 , 0.77)	0.64 (0.53 , 0.80)	0.62 (0.47 , 0.82)	0.68 (0.50 , 0.86)

Table 26.1 Hazard ratios (and 95% CI) to compare strategy of "treat within the 3 intervals following that when HbA1c exceeds x%" for X = 8, 9, 10 and reference strategy of x=7 for reaching target HbA1c.

[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).

	Unweighted model baseline adjusted – hazard ratio for strategy vs 6.5% for risk MI: three 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%	0.89 (0.84 , 0.95)	1.01 (0.86 , 1.22)	1.03 (0.89 , 1.20)	1.20 (1.01 , 1.54)	1.08 (0.83 , 1.51)	1.00 (0.89 , 1.14)
8%	0.91 (0.83 , 1.01)	0.95 (0.80 , 1.18)	1.14 (0.96 , 1.38)	1.17 (0.97 , 1.50)	1.17 (0.89 , 1.62)	1.01 (0.88 , 1.20)
9%	0.94 (0.84 , 1.06)	0.96 (0.78 , 1.25)	1.15 (0.95 , 1.41)	1.21 (0.98 , 1.58)	1.20 (0.91 , 1.77)	1.01 (0.88 , 1.19)
10%	0.98 (0.88 , 1.10)	0.93 (0.72 , 1.23)	1.18 (0.95 , 1.45)	1.22 (0.97 , 1.59)	1.24 (0.91 , 1.85)	1.05 (0.90 , 1.26)
	IPW* Dynamic MSM – hazard ratio for strategy vs 6.5% for risk MI: three 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%	0.94 (0.88 , 1.01)	1.07 (0.93 , 1.38)	1.11 (0.97 , 1.33)	1.31 (1.04 , 1.77)	1.00 (0.53 , 2.24)	0.97 (0.81 , 1.23)
8%	1.04 (0.94 , 1.18)	0.96 (0.57 , 1.60)	1.42 (1.15 , 1.88)	1.34 (0.87 , 2.07)	1.07 (0.61 , 2.59)	0.96 (0.70 , 1.38)
9%	1.14 (0.99 , 1.33)	1.00 (0.56 , 1.87)	1.44 (1.07 , 2.15)	1.59 (0.91 , 2.74)	1.07 (0.58 , 2.90)	0.90 (0.66 , 1.30)
10%	1.21 (1.03 , 1.42)	0.95 (0.51 , 1.88)	1.50 (1.06 , 2.38)	1.66 (0.95 , 2.80)	1.24 (0.60 , 3.51)	0.98 (0.66 , 1.53)

Table 26.2 Hazard ratios (and 95% CI) to compare strategy of "treat within the 3 intervals following that when HbA1c exceeds x%" for X = 8, 9, 10 and reference strategy of x=7 for risk of MI.

[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).

	Unweighted model baseline adjusted – hazard ratio for strategy vs 6.5% for risk of stroke: three 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%	1.06 (1.00 , 1.14)	0.87 (0.77 , 0.97)	1.08 (0.97 , 1.21)	0.96 (0.85 , 1.10)	0.98 (0.85 , 1.11)	0.99 (0.90 , 1.10)
8%	1.05 (0.97 , 1.16)	0.88 (0.74 , 1.03)	1.05 (0.94 , 1.21)	0.87 (0.75 , 1.04)	0.88 (0.73 , 1.05)	1.03 (0.90 , 1.15)
9%	1.07 (0.97 , 1.17)	0.91 (0.77 , 1.08)	1.03 (0.91 , 1.19)	0.91 (0.78 , 1.09)	0.85 (0.69 , 1.02)	1.00 (0.85 , 1.14)
10%	1.06 (0.97 , 1.19)	0.89 (0.75 , 1.09)	1.04 (0.93 , 1.21)	0.89 (0.74 , 1.07)	0.88 (0.71 , 1.10)	1.00 (0.88 , 1.15)
	IPW* Dynamic MSM – hazard ratio for strategy vs 6.5% for risk of stroke: 3 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%	1.05 (0.97 , 1.14)	0.94 (0.85 , 1.10)	0.88 (0.75 , 1.13)	1.01 (0.90 , 1.19)	0.99 (0.86 , 1.19)	0.99 (0.78 , 1.27)
8%	0.98 (0.77 , 1.20)	0.85 (0.56 , 1.26)	0.71 (0.52 , 1.08)	0.96 (0.68 , 1.39)	0.91 (0.55 , 1.53)	1.19 (0.88 , 1.67)
9%	0.97 (0.71 , 1.24)	0.79 (0.49 , 1.29)	0.68 (0.49 , 1.09)	0.96 (0.63 , 1.41)	0.87 (0.52 , 1.51)	1.11 (0.74 , 1.58)
10%	0.92 (0.70 , 1.21)	0.79 (0.49 , 1.32)	0.69 (0.50 , 1.11)	0.88 (0.55 , 1.30)	0.91 (0.53 , 1.57)	1.04 (0.71 , 1.57)

Table 26.3 Hazard ratios (and 95% CI) to compare risk of stroke between strategy of "treat within 3 intervals 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).

	Unweighted model baseline adjusted – hazard ratio for strategy vs 6.5% for risk of all-cause mortality: three 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%	1.02 (0.97 , 1.07)	0.94 (0.87 , 1.01)	1.01 (0.95 , 1.08)	0.97 (0.91 , 1.05)	0.98 (0.91 , 1.08)	1.02 (0.98 , 1.06)
8%	1.03 (0.96 , 1.10)	0.93 (0.84 , 1.05)	0.98 (0.92 , 1.07)	0.97 (0.89 , 1.06)	0.98 (0.88 , 1.10)	1.06 (1.01 , 1.12)
9%	1.05 (0.99 , 1.13)	0.93 (0.84 , 1.05)	0.95 (0.88 , 1.04)	1.00 (0.90 , 1.09)	0.97 (0.86 , 1.09)	1.07 (1.02 , 1.15)
10%	1.06 (0.99 , 1.15)	0.92 (0.83 , 1.05)	0.97 (0.89 , 1.05)	1.00 (0.90 , 1.11)	0.98 (0.87 , 1.11)	1.08 (1.02 , 1.16)
	IPW* dynamic MSM – hazard ratio for strategy vs 6.5% for risk of all-cause mortality: three 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%	1.10 (1.05 , 1.16)	1.05 (0.93 , 1.30)	0.98 (0.85 , 1.10)	1.02 (0.90 , 1.20)	0.97 (0.86 , 1.13)	1.07 (0.99 , 1.17)
8%	1.22 (1.13 , 1.32)	0.92 (0.71 , 1.34)	1.04 (0.84 , 1.27)	1.05 (0.87 , 1.34)	0.98 (0.77 , 1.26)	1.11 (0.96 , 1.29)
9%	1.31 (1.21 , 1.43)	0.97 (0.75 , 1.38)	0.89 (0.73 , 1.15)	1.09 (0.87 , 1.51)	0.99 (0.76 , 1.30)	1.15 (0.98 , 1.36)
10%	1.34 (1.24 , 1.46)	0.98 (0.77 , 1.41)	0.92 (0.73 , 1.20)	1.09 (0.87 , 1.45)	1.07 (0.81 , 1.38)	1.13 (0.95 , 1.34)

Table 26.4 Hazard ratios (and 95% CI) to compare risk of all-cause mortality between strategy of "treat within 3 intervals 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).

APPENDIX 27 FULL RESULTS OF CHAPTER 9 SENSITIVITY ANALYSES – INITIATION WITH METFORMIN ONLY IN THE DYNAMIC STRATEGY.

HbA1c

	IPW* & IPCW dynamic MSM – hazard ratio for strategy vs 7% for risk of all-cause mortality: one month grace period (HR<1 indicates superior strategy): ITT					
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%	0.87 (0.83 , 0.93)	0.79 (0.74 , 0.87)	0.80 (0.72 , 0.87)	0.87 (0.75 , 1.03)	1.00 (0.79 , 1.21)	0.98 (0.80 , 1.2)
9%	0.86 (0.82 , 0.92)	0.74 (0.68 , 0.83)	0.72 (0.64 , 0.8)	0.75 (0.60 , 0.89)	0.83 (0.63 , 1.08)	0.81 (0.64 , 1.03)
10%	0.86 (0.81 , 0.92)	0.74 (0.68 , 0.82)	0.68 (0.61 , 0.76)	0.68 (0.57 , 0.82)	0.65 (0.48 , 0.83)	0.64 (0.48 , 0.84)
	IPW* & IPCW dynamic MSM – hazard ratio for strategy vs 7% for risk of all-cause mortality: one month grace period (HR<1 indicates superior strategy: As treated					
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%	0.87 (0.82 , 0.92)	0.79 (0.73 , 0.86)	0.77 (0.71 , 0.87)	0.82 (0.69 , 0.99)	1.00 (0.71 , 1.41)	1.01 (0.76 , 1.35)
9%	0.86 (0.81 , 0.90)	0.75 (0.69 , 0.81)	0.69 (0.63 , 0.79)	0.71 (0.57 , 0.83)	0.73 (0.5 , 1.04)	0.88 (0.64 , 1.32)
10%	0.86 (0.81 , 0.91)	0.74 (0.69 , 0.80)	0.66 (0.61 , 0.75)	0.66 (0.52 , 0.77)	0.71 (0.47 , 1.00)	0.65 (0.45 , 0.95)

Table 27.1 Hazard ratios (and 95% CI) to compare risk of target attainment between strategy of "treat with metformin in the intervals following that when HbA1c exceeds x%" for X = 8, 9, 10 and strategy of x=7. presented by time since study entry.

Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW).

[1]Columns show estimated HR and 95% CI by time since study entry. CI's obtained via 200 bootstrap replications

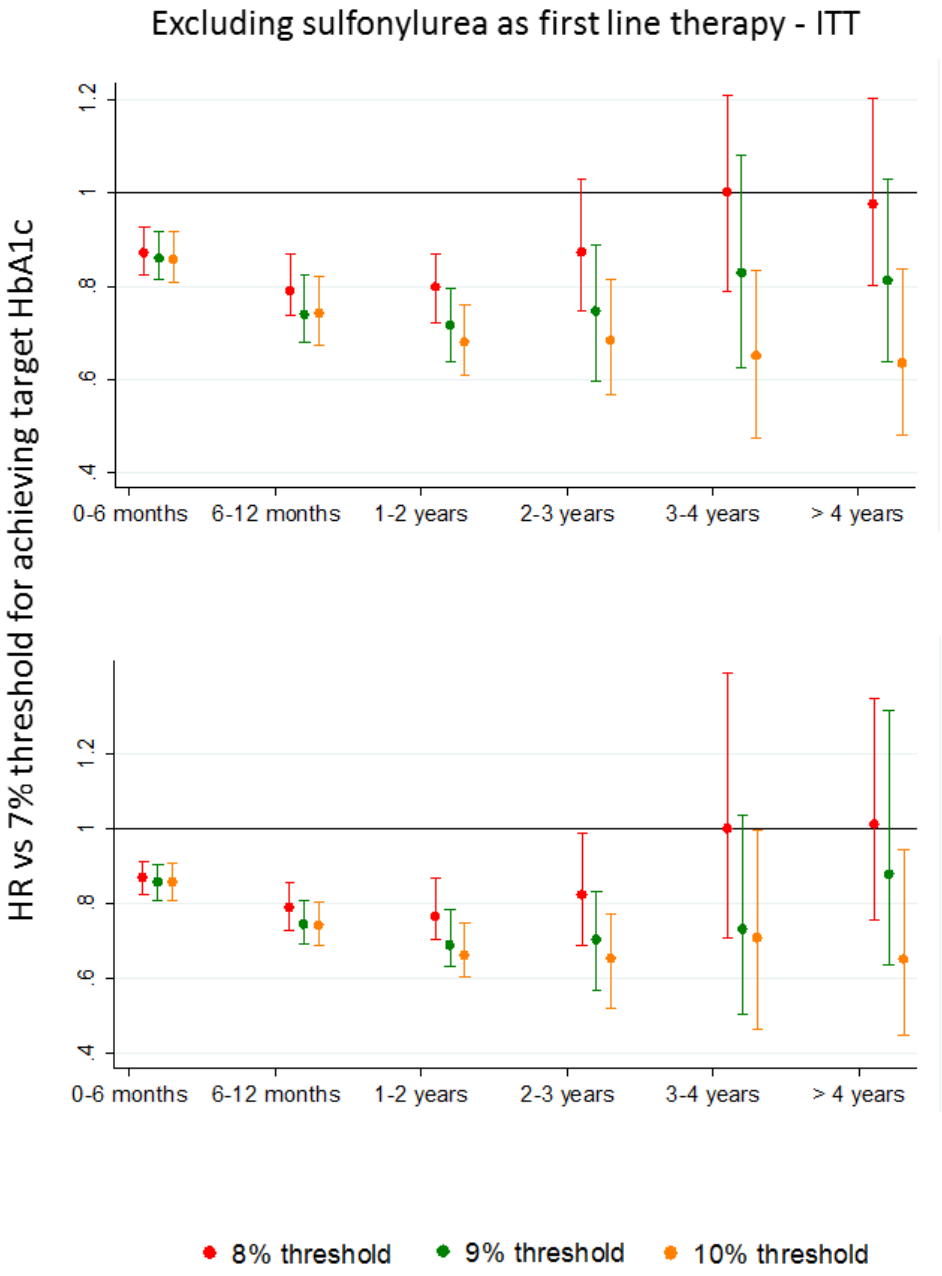
* weighting model includes: age gender, calendar period of diabetes onset (pre or post 2005), smoking status, alcohol consumption, use in the previous year of anti-hypertensive drugs, statins, NSAIDS or aspirin (baseline and time updated); previous history of any CVD, stroke, MI, CKD or cancer (baseline and time updated); HbA1c, BMI and SBP (baseline and time updated).

		IPW Dynamic MSM (one-month grace period): ITT approach Proportion achieving target HBA1c by…			IPW Dynamic MSM (one-month grace period): As treated approach Proportion achieving target HBA1c by…		
		1 year	**2 years**	**4 years**	**1 year**	**2 years**	**4 years**
HbA1c threshold	7%	0.36 (0.35 , 0.37)	0.49 (0.47 , 0.50)	0.59 (0.58 , 0.61)	0.36 (0.35 , 0.37)	0.49 (0.47 , 0.50)	0.60 (0.57 , 0.61)
	8%	0.31 (0.31 , 0.32)	0.42 (0.41 , 0.43)	0.54 (0.53 , 0.55)	0.31 (0.30 , 0.32)	0.42 (0.41 , 0.43)	0.53 (0.52 , 0.54)
	9%	0.31 (0.30 , 0.31)	0.41 (0.40 , 0.41)	0.51 (0.49 , 0.52)	0.30 (0.30 , 0.31)	0.4 (0.40 , 0.41)	0.50 (0.48 , 0.51)
	10%	0.31 (0.30 , 0.31)	0.40 (0.39 , 0.41)	0.49 (0.48 , 0.50)	0.30 (0.30 , 0.31)	0.40 0.39 , 0.41)	0.49 (0.48 , 0.5)

Table 27.2 Estimated proportions of population achieving target HbA1c by 1, 2 and 4 years from study entry, for each treatment strategy.

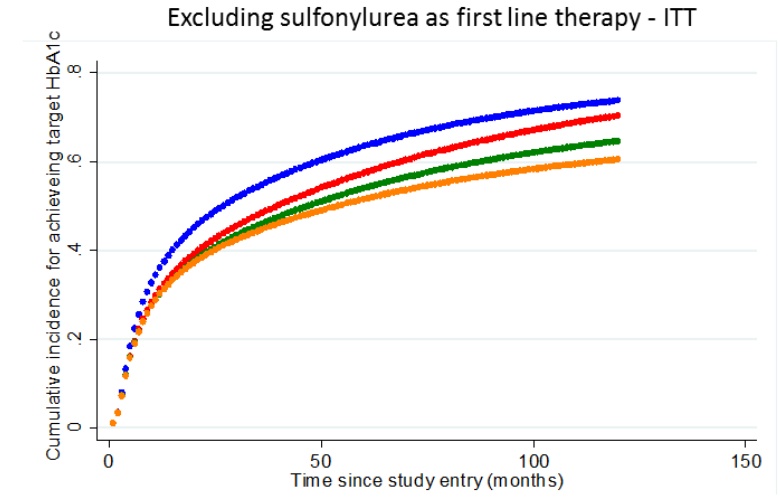
CI's obtained via 200 bootstrap replications

Figure 27.1 Estimated hazard ratios (and 95% CI) comparing strategy of "treat with metformin in the interval following that when HbA1c first exceeds x%" for X = 8, 9, 10 and reference strategy of x=7. for risk of reaching target HbA1c, presented by time since study entry.

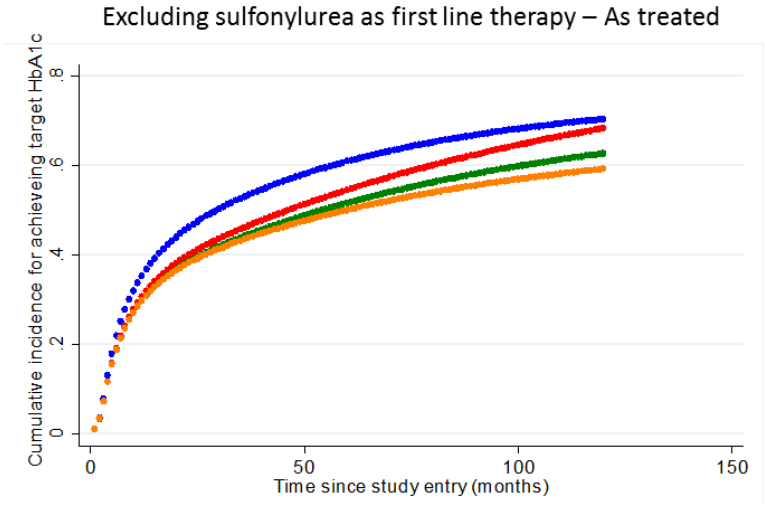


Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW). CI's obtained via 200 bootstrap replications.

Figure 27.2 Estimated cumulative incidence curves for achieving target HbA1c of 6.5% for different thresholds for metformin initiation.



Month		0	20	40	60	80	100	120
	7	31420	3932	1628	864	486	221	94
Number still	8	31420	6457	2474	1133	599	276	121
at risk	9	31420	6481	2277	961	481	214	96
	10	31420	6249	2052	815	390	159	71



Month		0	20	40	60	80	100	120
	7	31309	3393	1082	420	164	46	10
Number still	8	31309	5938	1922	679	259	92	32
at risk	9	31309	6065	1856	611	230	72	28
	10	31309	5953	1757	576	218	70	23

Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW).

MI

	IPW* & IPCW Dynamic MSM – hazard ratio for strategy vs 6.5% for risk MI: one month grace period (HR<1 indicates superior strategy): ITT					
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%	0.72 (0.44 , 1.06)	1.20 (1.02 , 1.64)	1.24 (1.05 , 1.77)	0.83 (0.38 , 2.41)	1.16 (0.64 , 2.56)	0.98 (0.70 , 1.41)
8%	0.83 (0.55 , 1.24)	1.49 (1.23 , 2.14)	1.49 (0.90 , 2.37)	1.49 (0.68 , 4.90)	1.25 (0.47 , 3.9)	1.22 (0.82 , 2.02)
9%	0.84 (0.54 , 1.28)	1.70 (1.31 , 2.76)	1.50 (0.91 , 2.57)	1.89 (0.80 , 5.88)	1.17 (0.49 , 3.38)	1.07 (0.69 , 1.76)
10%	0.90 (0.57 , 1.39)	1.64 (1.14 , 2.91)	1.66 (0.98 , 2.85)	1.74 (0.75 , 5.11)	1.19 (0.58 , 3.41)	1.25 (0.77 , 2.07)
	IPW* & IPCW Dynamic MSM – hazard ratio for strategy vs 6.5% for risk MI: one month grace period (HR<1 indicates superior strategy): As treated					
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%	0.80 (0.47 , 1.36)	1.30 (1.07 , 1.59)	1.22 (0.97 , 1.53)	0.69 (0.31 , 1.54)	1.25 (0.64 , 2.46)	0.92 (0.60 , 1.42)
8%	0.92 (0.52 , 1.65)	1.66 (1.30 , 2.11)	1.38 (0.82 , 2.33)	1.22 (0.51 , 2.94)	1.75 (0.57 , 5.34)	1.22 (0.70 , 2.11)
9%	0.91 (0.50 , 1.66)	1.84 (1.33 , 2.54)	1.47 (0.81 , 2.67)	1.61 (0.65 , 3.96)	1.48 (0.49 , 4.54)	1.19 (0.68 , 2.10)
10%	0.98 (0.54 , 1.80)	1.91 (1.30 , 2.81)	1.51 (0.75 , 3.02)	1.40 (0.57 , 3.45)	1.64 (0.54 , 5.01)	1.34 (0.76 , 2.37)

Table 27.3 Hazard ratios (and 95% CI)to compare risk of MI between strategy of "treat with metformin in the intervals following that when HbA1c exceeds x%" for X = 7, 8, 9, 10 and strategy of x=6.5. presented by time since study entry

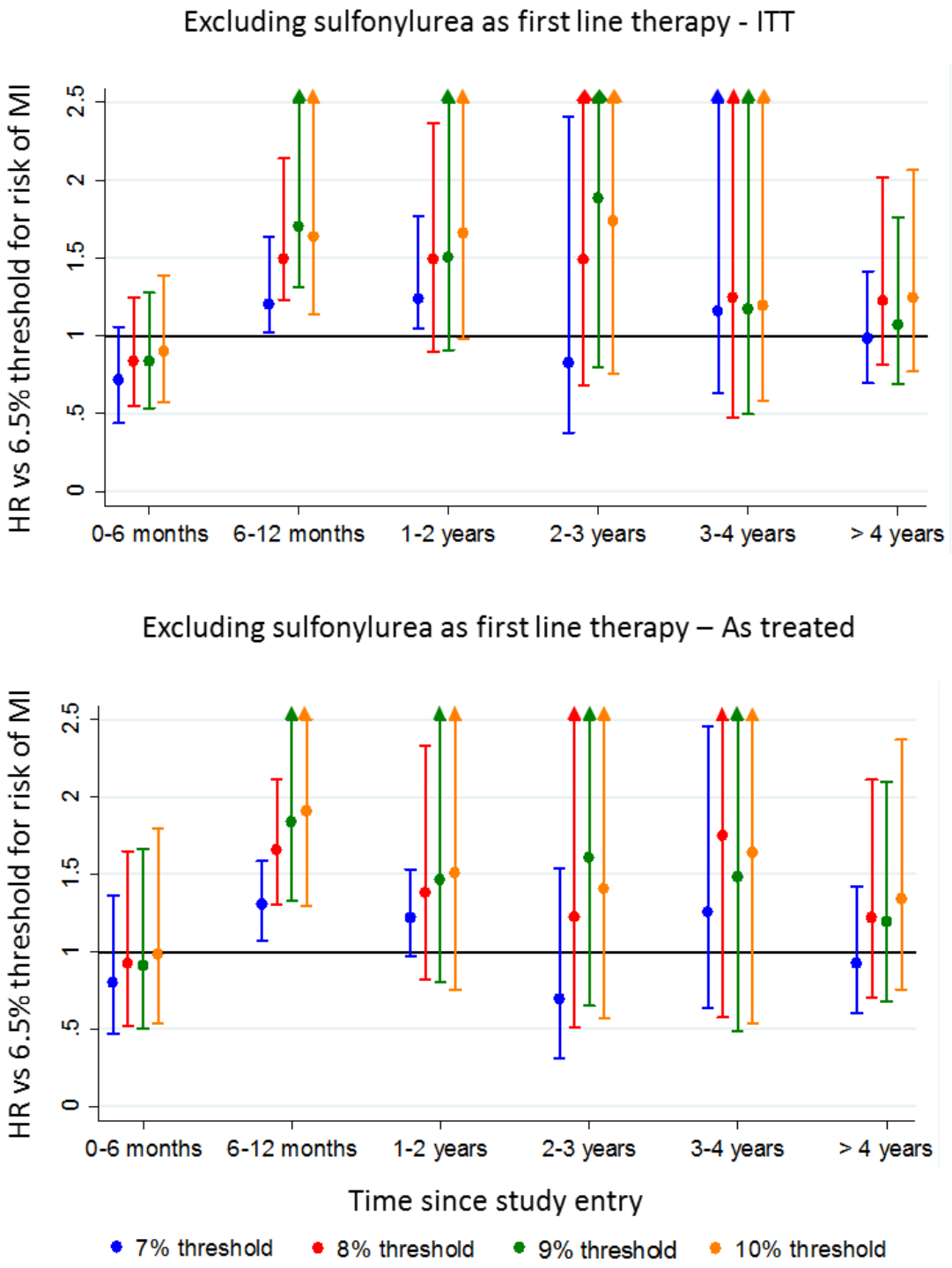
Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW).[1]Columns show estimated HR 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).

		IPW & IPCW Dynamic MSM (one-month grace period): ITT approach Percent experiencing MI by…			IPW & IPCW Dynamic MSM (one-month grace period): As treated approach Percent experiencing MI by…		
		1 year	**2 years**	**4 years**	**1 year**	**2 years**	**4 years**
HbA1c threshold	6.5%	0.39 (0.27 , 0.55)	0.62 (0.44 , 0.84)	1.11 (0.83 , 1.54)	0.34 (0.22 , 0.50)	0.60 (0.42 , 0.87)	1.12 (0.79 , 1.60)
	7%	0.35 (0.26 , 0.44)	0.64 (0.50 , 0.79)	1.13 (0.88 , 1.41)	0.33 (0.25 , 0.45)	0.65 (0.51 , 0.83)	1.14 (0.88 , 1.44)
	8%	0.41 (0.33 , 0.50)	0.76 (0.64 , 0.88)	1.41 (1.16 , 1.67)	0.40 (0.31 , 0.50)	0.76 (0.63 , 0.93)	1.51 (1.22 , 1.89)
	9%	0.44 (0.35 , 0.53)	0.79 (0.66 , 0.93)	1.50 (1.23 , 1.80)	0.41 (0.33 , 0.53)	0.80 (0.65 , 0.99)	1.61 (1.31 , 2.09)
	10%	0.44 (0.36 , 0.53)	0.83 (0.71 , 0.96)	1.52 (1.27 , 1.81)	0.44 (0.34 , 0.56)	0.83 (0.68 , 1.01)	1.62 (1.32 , 2.01)

Table 27.4 Estimated percentage of population experiencing an MI by 1, 2 and 4 years from study entry, for each treatment strategy.

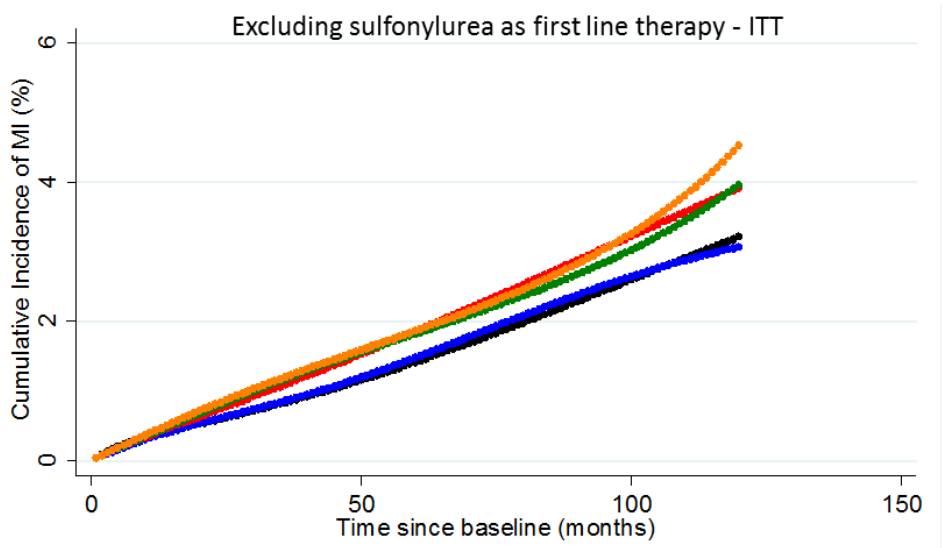
CI's obtained via 200 bootstrap replications.

Figure 27.3 Estimated hazard ratios (and 95% CIs) comparing strategy of "treat with metformin in the interval following that when HbA1c first exceeds x%" for X = 8, 9, 10 and reference strategy of x=7. for risk of MI, presented by time since study entry.

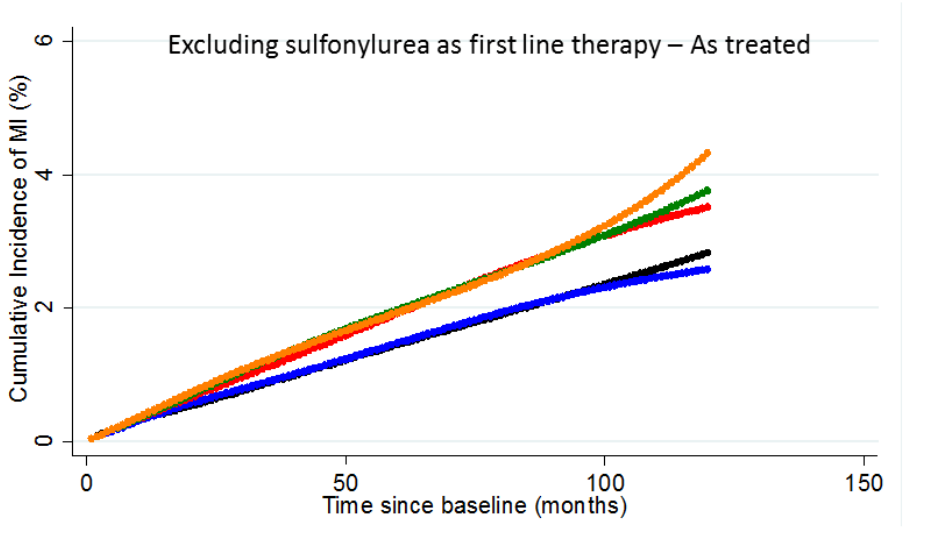


Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW). CI's obtained via 200 bootstrap replications.

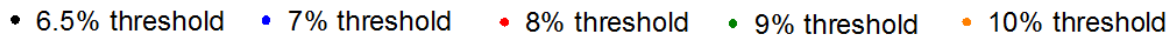
Figure 27.4 Estimated cumulative incidence of MI (%) for different thresholds for metformin initiation.



Month		0	20	40	60	80	100	120
Number still at risk	6.5	49061	14098	8723	5381	3186	1752	905
	7	49061	21715	13557	8477	5088	2821	1458
	8	49061	25952	15962	9878	5871	3267	1668
	9	49061	26073	15552	9358	5430	2951	1465
	10	49061	14098	8723	5381	3186	1752	905



Month		0	20	40	60	80	100	120
Number still at risk	6.5	48949	13351	7846	4548	2511	1293	642
	7	48949	20936	12565	7484	4240	2231	1097
	8	48949	25233	15036	8953	5061	2695	1323
	9	48949	25505	14867	8676	4858	2565	1245
	10	48949	25395	14663	8533	4773	2517	1219



Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW).

Stroke

	IPW *& IPCW Dynamic MSM – hazard ratio for strategy vs 6.5% for risk of stroke: one month grace period (HR<1 indicates superior strategy) : ITT					
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 (1.07 , 1.37)	0.77 (0.50 , 1.14)	1.52 (1.20 , 2.12)	0.79 (0.63 , 1.00)	1.18 (0.85 , 1.72)	0.86 (0.67 , 1.15)
8%	1.27 (0.92 , 1.68)	0.77 (0.48 , 1.25)	1.87 (1.22 , 2.99)	0.64 (0.43 , 1.01)	0.81 (0.53 , 1.28)	1.04 (0.79 , 1.43)
9%	1.37 (0.98 , 1.95)	0.87 (0.55 , 1.43)	1.88 (1.22 , 2.79)	0.69 (0.45 , 1.13)	0.67 (0.41 , 1.23)	1.02 (0.78 , 1.4)
10%	1.26 (0.90 , 1.86)	0.82 (0.51 , 1.38)	1.93 (1.25 , 2.94)	0.58 (0.37 , 0.97)	0.79 (0.48 , 1.56)	0.95 (0.74 , 1.33)
	IPW *& IPCW Dynamic MSM – hazard ratio for strategy vs 6.5% for risk of stroke: one month grace period (HR<1 indicates superior strategy): AS treated					
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%	1.16 (1.08 , 1.29)	0.78 (0.48 , 1.25)	1.45 (1.19 , 1.92)	0.84 (0.68 , 1.00)	1.32 (0.84 , 2.13)	0.99 (0.69 , 1.47)
8%	1.20 (0.90 , 1.57)	0.90 (0.57 , 1.46)	1.69 (1.13 , 2.67)	0.64 (0.44 , 1.06)	0.94 (0.54 , 1.87)	1.34 (0.95 , 2.05)
9%	1.31 (0.95 , 1.94)	0.97 (0.63 , 1.56)	1.68 (1.14 , 2.60)	0.63 (0.41 , 1.09)	0.71 (0.42 , 1.51)	1.24 (0.83 , 2.02)
10%	1.18 (0.87 , 1.72)	0.92 (0.55 , 1.59)	1.79 (1.22 , 2.74)	0.61 (0.40 , 1.03)	0.80 (0.47 , 1.70)	1.30 (0.89 , 2.19)

Table 27.5 Hazard ratios (and 95% CIs) to compare risk of stroke between strategy of "treat with metformin in the intervals following that when HbA1c exceeds x%" for X = 7, 8, 9, 10 and strategy of x=6.5. presented by time since study entry.

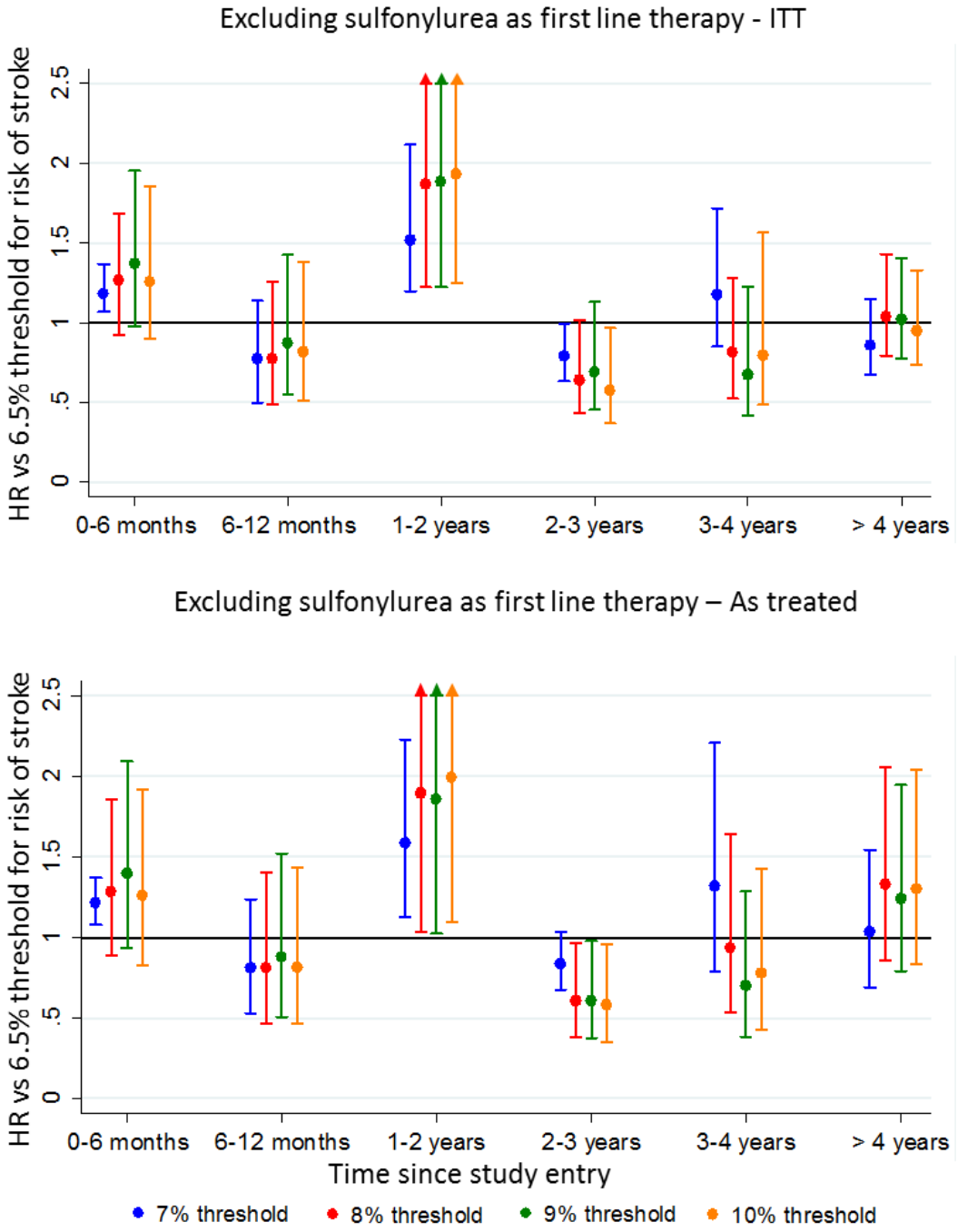
Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW).[1]Columns show estimated HR 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).

		IPW & IPCW Dynamic MSM (one-month grace period): ITT approach Percent experiencing stroke by…			IPW & IPCW Dynamic MSM (one-month grace period): As treated approach Percent experiencing stroke by…		
HbA1c threshold		**1 year**	**2 years**	**4 years**	**1 year**	**2 years**	**4 years**
	6.5%	0.57 (0.43 , 0.74)	0.83 (0.64 , 1.05)	2.36 (1.89 , 2.89)	0.57 (0.43 , 0.74)	0.83 (0.64 , 1.05)	2.36 (1.89 , 2.89)
	7%	0.53 (0.41 , 0.68)	0.93 (0.76 , 1.11)	2.41 (1.96 , 2.89)	0.53 (0.41 , 0.68)	0.93 (0.76 , 1.11)	2.41 (1.96 , 2.89)
	8%	0.55 (0.46 , 0.66)	1.04 (0.91 , 1.18)	2.14 (1.92 , 2.41)	0.55 (0.46 , 0.66)	1.04 (0.91 , 1.18)	2.14 (1.92 , 2.41)
	9%	0.60 (0.50 , 0.71)	1.10 (0.95 , 1.26)	2.14 (1.91 , 2.39)	0.60 (0.50 , 0.71)	1.10 (0.95 , 1.26)	2.14 (1.91 , 2.39)
	10%	0.56 (0.46 , 0.66)	1.07 (0.93 , 1.22)	2.10 (1.88 , 2.38)	0.56 (0.46 , 0.66)	1.07 (0.93 , 1.22)	2.10 (1.88 , 2.38)

Table 27.6 Estimated percentage of population experiencing a stroke by 1, 2 and 4 years from study entry, for each treatment strategy

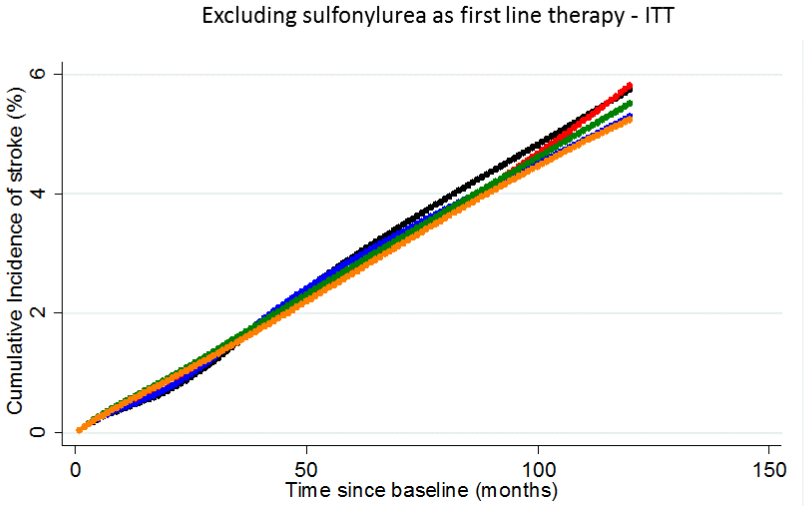
CI's obtained via 200 bootstrap replications.

Figure 27.5 Estimated hazard ratios (and 95% CIs) comparing strategy of "treat with metformin in the interval following that when HbA1c first exceeds x%" for X = 8, 9, 10 and reference strategy of x=7. for risk of stroke, presented by time since study entry.

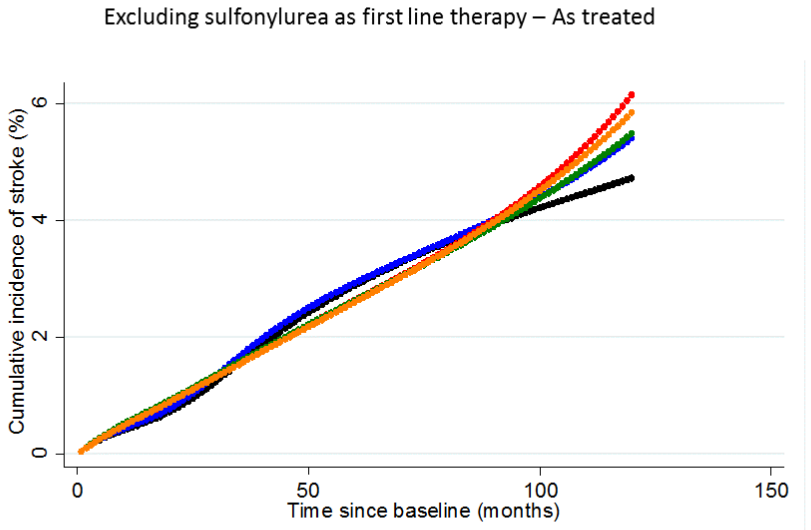


Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW). CI's obtained via 200 bootstrap replications.

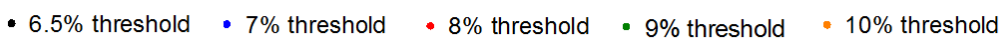
Figure 27.6 Estimated cumulative incidence (%) curves for risk of stroke, for different thresholds for metformin initiation.



Month		0	20	40	60	80	100	120
Number still at risk	6.5	49023	14049	8629	5304	3132	1713	889
	7	49023	21627	13407	8364	5008	2750	1417
	8	49023	25844	15803	9762	5791	3193	1632
	9	49023	25969	15395	9241	5347	2890	1431
	10	49023	25701	14984	8874	5063	2707	1336



Month		0	20	40	60	80	100	120
Number still at risk	6.5	48911	13302	7754	4479	2460	1260	631
	7	48911	20849	12420	7383	4169	2171	1070
	8	48911	25128	14879	8843	4982	2628	1294
	9	48911	25403	14710	8567	4778	2500	1211
	10	48911	25293	14508	8428	4696	2456	1187



Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW).

All-cause mortality

	IPW* & IPCW dynamic MSM – hazard ratio for strategy vs 6.5% for risk of all-cause mortality: one month grace period (HR<1 indicates superior strategy: ITT					
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%	1.12 (0.94 , 1.31)	0.96 (0.80 , 1.13)	1.12 (0.99 , 1.31)	1.06 (0.88 , 1.25)	0.91 (0.75 , 1.09)	0.94 (0.85 , 1.04)
8%	1.18 (0.92 , 1.56)	0.94 (0.69 , 1.25)	1.05 (0.87 , 1.3)	1.10 (0.88 , 1.40)	0.85 (0.64 , 1.14)	1.05 (0.92 , 1.19)
9%	1.24 (0.95 , 1.67)	0.89 (0.66 , 1.20)	0.97 (0.79 , 1.27)	1.24 (0.96 , 1.61)	0.83 (0.61 , 1.14)	1.05 (0.90 , 1.2)
10%	1.25 (0.96 , 1.75)	0.89 (0.65 , 1.21)	1.00 (0.82 , 1.31)	1.30 (0.96 , 1.68)	0.90 (0.63 , 1.26)	1.06 (0.90 , 1.21)
IPW* & IPCW dynamic MSM – hazard ratio for strategy vs 6.5% for risk of all-cause mortality: one month grace period (HR<1 indicates superior strategy: As treated						
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%	1.13 (0.91 , 1.41)	0.95 (0.79 , 1.15)	1.15 (0.96 , 1.38)	1.06 (0.82 , 1.37)	0.89 (0.71 , 1.12)	0.90 (0.78 , 1.04)
8%	1.22 (0.89 , 1.68)	0.94 (0.69 , 1.29)	1.03 (0.79 , 1.33)	1.10 (0.83 , 1.45)	0.80 (0.59 , 1.08)	1.03 (0.86 , 1.24)
9%	1.28 (0.91 , 1.78)	0.89 (0.63 , 1.25)	0.97 (0.74 , 1.29)	1.22 (0.90 , 1.64)	0.77 (0.56 , 1.06)	1.02 (0.84 , 1.23)
10%	1.33 (0.95 , 1.87)	0.89 (0.63 , 1.26)	1.01 (0.77 , 1.34)	1.21 (0.87 , 1.67)	0.87 (0.62 , 1.21)	1.03 (0.85 , 1.26)

Table 27.7 Hazard ratios (and 95% CI) to compare risk of all-cause mortality between strategy of "treat with metformin in the intervals following that when HbA1c exceeds x%" for X = 7, 8, 9, 10 and strategy of x=6.5. presented by time since study entry

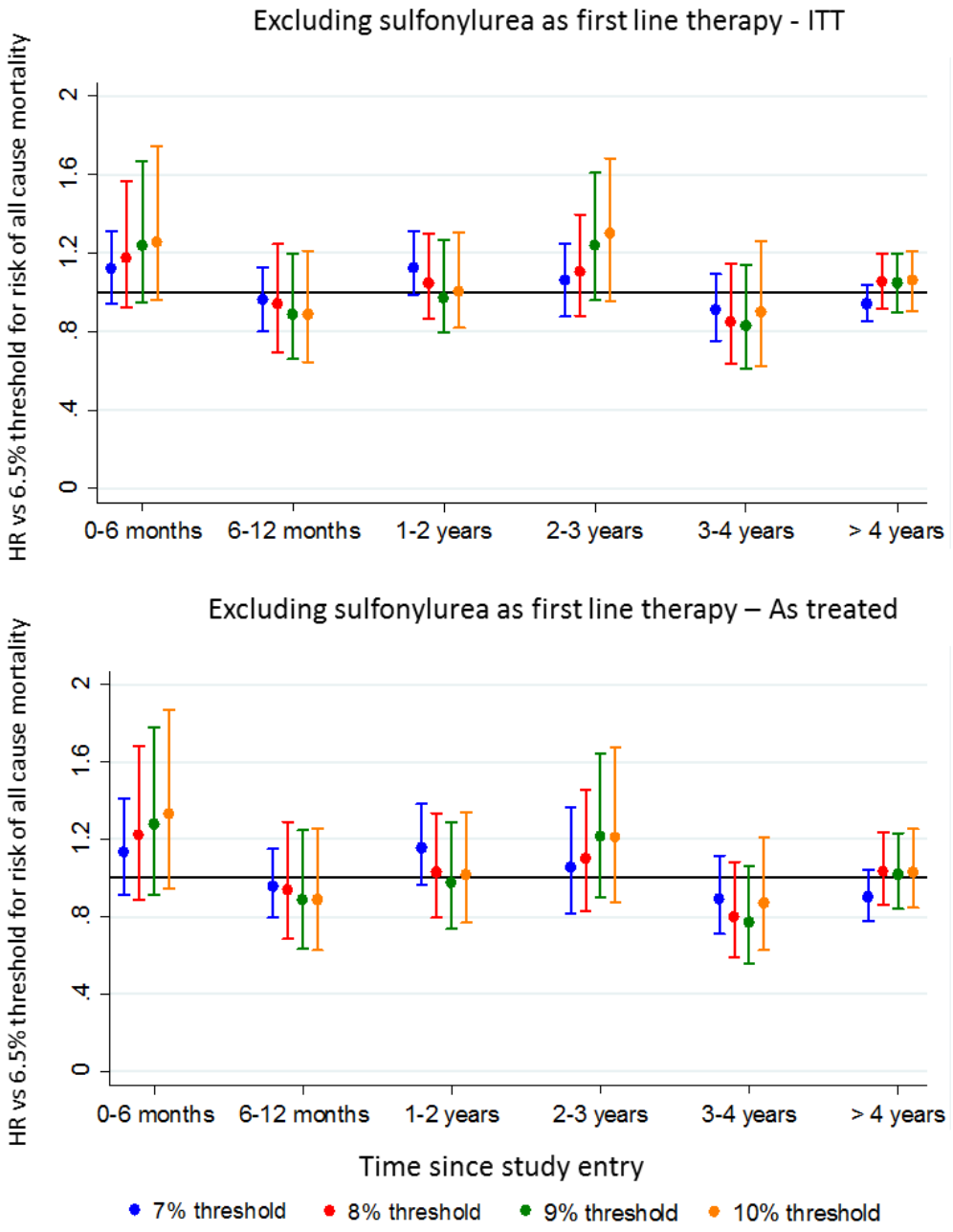
Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW).[1]Columns show estimated HR 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).

		IPW & IPCW Dynamic MSM (one-month grace period): ITT approach Percent mortality by…			IPW & IPCW Dynamic MSM (one-month grace period): As treated approach Percent mortality MI by…		
HbA1c threshold		**1 year**	**2 years**	**4 years**	**1 year**	**2 years**	**4 years**
	6.5%	1.41 (1.06 , 1.75)	2.85 (2.41 , 3.33)	6.65 (5.86 , 7.54)	1.43 (1.11 , 1.83)	2.83 (2.34 , 3.39)	6.75 (5.8 , 7.77)
	7%	1.44 (1.19 , 1.69)	3.06 (2.68 , 3.42)	6.73 (6.08 , 7.28)	1.45 (1.19 , 1.75)	3.06 (2.63 , 3.55)	6.79 (5.97 , 7.53)
	8%	1.45 (1.30 , 1.61)	2.96 (2.72 , 3.19)	6.56 (6.09 , 6.99)	1.47 (1.32 , 1.68)	2.91 (2.64 , 3.22)	6.50 (5.89 , 7.03)
	9%	1.43 (1.30 , 1.57)	2.83 (2.59 , 3.06)	6.58 (6.15 , 6.99)	1.45 (1.29 , 1.67)	2.81 (2.51 , 3.11)	6.51 (5.91 , 7.04)
	10%	1.44 (1.30 , 1.57)	2.89 (2.67 , 3.10)	6.89 (6.45 , 7.39)	1.48 (1.32 , 1.67)	2.89 (2.62 , 3.18)	6.81 (6.21 , 7.40)

Table 27.8 Estimated percentage of all-cause mortality by 1, 2 and 4 years from study entry, for each treatment strategy.

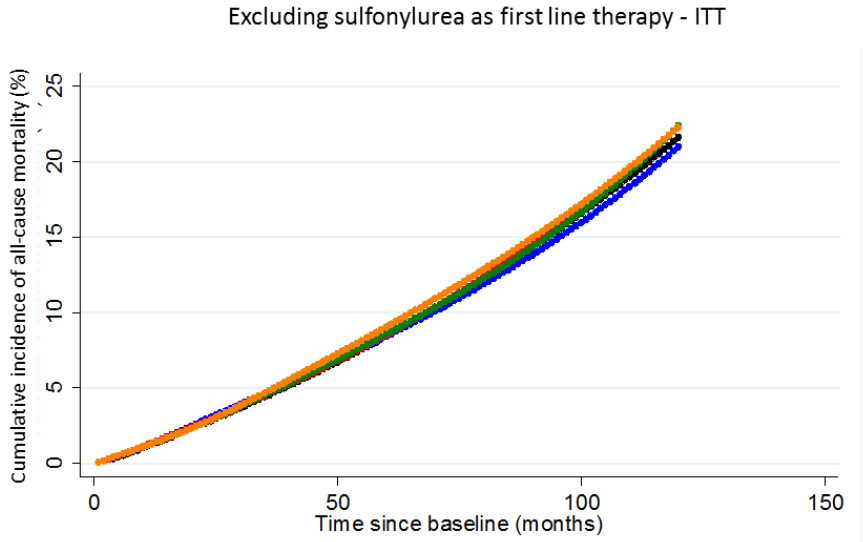
CI's obtained via 200 bootstrap replications

Figure 27.7 Estimated hazard ratios (and 95% CIs) comparing strategy of "treat with metformin in the interval following that when HbA1c first exceeds x%" for X = 8, 9, 10 and reference strategy of x=7. for risk of all-cause mortality, presented by time since study entry.

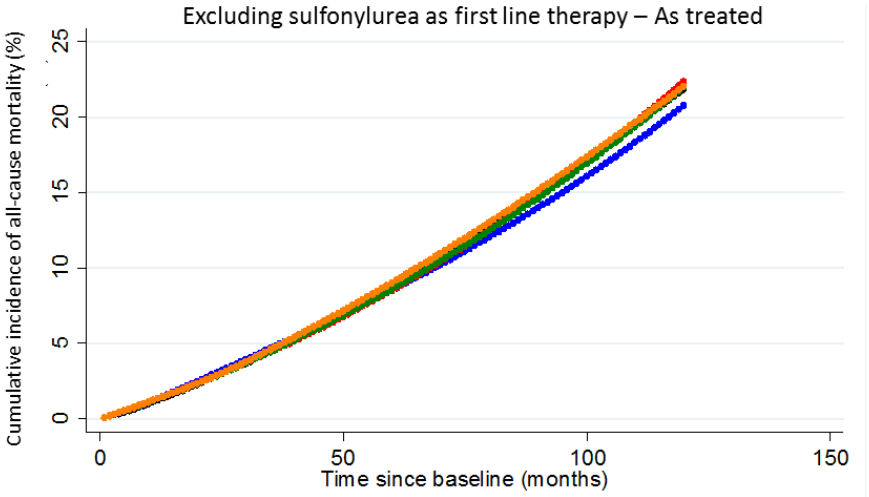


Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW). CI's obtained via 200 bootstrap replications.

Figure 27.8 Estimated cumulative incidence (%) curves for risk of all-cause mortality, for different thresholds for metformin initiation.



Month		0	20	40	60	80	100	120
Number still at risk	6.5	49149	14212	8820	5465	3246	1792	932
	7	49149	21882	13710	8627	5198	2884	1495
	8	49149	26160	16153	10069	6011	3352	1724
	9	49149	26290	15744	9544	5554	3028	1512
	10	49149	26016	15324	9162	5260	2838	1411



Month		0	20	40	60	80	100	120
Number still at risk	6.5	49037	13456	7930	4620	2558	1322	665
	7	49037	21095	12706	7619	4333	2281	1130
	8	49037	25433	15213	9124	5179	2765	1371
	9	49037	25715	15045	8848	4970	2630	1286
	10	49037	25604	14841	8704	4886	2582	1260

• 6.5% threshold • 7% threshold • 8% threshold • 9% threshold • 10% threshold

Results from model assuming ITT approach (top) and as treated approach (bottom). These models have joint IPW and IPCW to account for informative censoring by non-compliance to regime (as in primary analysis), and censoring due to sulfonylurea initiation (additional IPCW).