



Comparative effectiveness of second line oral antidiabetic treatments among people with type 2 diabetes mellitus: emulation of a target trial using routinely collected health data

Patrick Bidulka,¹ David G Lugo-Palacios,² Orlagh Carroll,² Stephen O'Neill,² Amanda I Adler,³ Anirban Basu,⁴ Richard J Silverwood,⁵ Jonathan W Bartlett,⁶ Dorothea Nitsch,¹ Paul Charlton,⁷ Andrew H Briggs,² Liam Smeeth,¹ Ian J Douglas,¹ Kamlesh Khunti,⁸ Richard Grieve²

For numbered affiliations see end of the article

Correspondence to: P Bidulka patrick.bidulka1@lshtm.ac.uk (or @pbidulka on X; ORCID 0000-0001-7644-2030)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2024;385:e077097 <http://dx.doi.org/10.1136/bmj-2023-077097>

Accepted: 14 March 2024

ABSTRACT OBJECTIVE

To compare the effectiveness of three commonly prescribed oral antidiabetic drugs added to metformin for people with type 2 diabetes mellitus requiring second line treatment in routine clinical practice.

DESIGN

Cohort study emulating a comparative effectiveness trial (target trial).

SETTING

Linked primary care, hospital, and death data in England, 2015-21.

PARTICIPANTS

75 739 adults with type 2 diabetes mellitus who initiated second line oral antidiabetic treatment with a sulfonylurea, DPP-4 inhibitor, or SGLT-2 inhibitor added to metformin.

MAIN OUTCOME MEASURES

Primary outcome was absolute change in glycated haemoglobin A_{1c} (HbA_{1c}) between baseline and one year follow-up. Secondary outcomes were change in body mass index (BMI), systolic blood pressure, and estimated glomerular filtration rate (eGFR) at one year

and two years, change in HbA_{1c} at two years, and time to ≥40% decline in eGFR, major adverse kidney event, hospital admission for heart failure, major adverse cardiovascular event (MACE), and all cause mortality. Instrumental variable analysis was used to reduce the risk of confounding due to unobserved baseline measures.

RESULTS

75 739 people initiated second line oral antidiabetic treatment with sulfonylureas (n=25 693, 33.9%), DPP-4 inhibitors (n=34 464, 45.5%), or SGLT-2 inhibitors (n=15 582, 20.6%). SGLT-2 inhibitors were more effective than DPP-4 inhibitors or sulfonylureas in reducing mean HbA_{1c} values between baseline and one year. After the instrumental variable analysis, the mean differences in HbA_{1c} change between baseline and one year were -2.5 mmol/mol (95% confidence interval (CI) -3.7 to -1.3) for SGLT-2 inhibitors versus sulfonylureas and -3.2 mmol/mol (-4.6 to -1.8) for SGLT-2 inhibitors versus DPP-4 inhibitors. SGLT-2 inhibitors were more effective than sulfonylureas or DPP-4 inhibitors in reducing BMI and systolic blood pressure. For some secondary endpoints, evidence for SGLT-2 inhibitors being more effective was lacking—the hazard ratio for MACE, for example, was 0.99 (95% CI 0.61 to 1.62) versus sulfonylureas and 0.91 (0.51 to 1.63) versus DPP-4 inhibitors. SGLT-2 inhibitors had reduced hazards of hospital admission for heart failure compared with DPP-4 inhibitors (0.32, 0.12 to 0.90) and sulfonylureas (0.46, 0.20 to 1.05). The hazard ratio for a ≥40% decline in eGFR indicated a protective effect versus sulfonylureas (0.42, 0.22 to 0.82), with high uncertainty in the estimated hazard ratio versus DPP-4 inhibitors (0.64, 0.29 to 1.43).

CONCLUSIONS

This emulation study of a target trial found that SGLT-2 inhibitors were more effective than sulfonylureas or DPP-4 inhibitors in lowering mean HbA_{1c}, BMI, and systolic blood pressure and in reducing the hazards of hospital admission for heart failure (v DPP-4 inhibitors) and kidney disease progression (v sulfonylureas), with no evidence of differences in other clinical endpoints.

Introduction

About 463 million people worldwide (9.3%) have type 2 diabetes mellitus.¹ In most people this disease is progressive, and it is associated with risks of multiple complications, including cardiovascular disease (CVD) and chronic kidney disease.² Interventions that improve biomarkers of type 2 diabetes mellitus,

WHAT IS ALREADY KNOWN ON THIS TOPIC

Placebo controlled randomised trials showed that sodium-glucose cotransporter-2 (SGLT-2) inhibitors are cardioprotective and kidney protective among people with type 2 diabetes mellitus (T2DM)

NICE guidelines recommend SGLT-2 inhibitors with metformin as second line oral antidiabetic treatment for people with T2DM and cardiovascular disease (CVD), or at high risk of CVD; however, for the broader population with T2DM without these indications, these guidelines recommend sulfonylureas, DPP-4 inhibitors, or SGLT-2 inhibitors along with metformin

The comparative effectiveness of these three second line treatments has not been assessed directly in randomised controlled trials, and evidence from observational studies is prone to confounding by indication

WHAT THIS STUDY ADDS

SGLT-2 inhibitors were found to be more effective than sulfonylureas or DPP-4 inhibitors in lowering mean haemoglobin A_{1c} levels, body mass index, and systolic blood pressure for a broad population of people with T2DM

SGLT-2 inhibitors were found to be more effective than sulfonylureas or DPP-4 inhibitors in reducing the hazards of hospital admission for heart failure (v DPP-4 inhibitors) and kidney disease progression (v sulfonylureas) for a broad population of people with T2DM

A target trial design was combined with an instrumental variable analysis to help reduce the risk of bias from confounding and to supplement previous studies in providing useful evidence that applies directly to routine clinical practice

such as glycated haemoglobin A_{1c} (HbA_{1c}), blood pressure, and lipid levels, can reduce the risk of these complications.³⁻⁶ International clinical guidelines recommend additional drugs (second line treatment) if glycaemic control is inadequate after metformin monotherapy.⁷⁻⁹ A recent study of second line treatments for people with type 2 diabetes mellitus across 38 countries reported that the most commonly used oral drugs were dipeptidyl peptidase-4 (DPP-4) inhibitors (48.3%), sulfonylureas (40.9%), and sodium-glucose cotransporter-2 (SGLT-2) inhibitors (8.3%).¹⁰

Of these oral treatments, SGLT-2 inhibitors are newer and more costly classes of drugs.¹¹ In England, SGLT-2 inhibitors are recommended second line treatments in preference to other drug classes for some people with type 2 diabetes mellitus—those with pre-existing CVD, at high risk of CVD, or with kidney disease.⁷ For most people with type 2 diabetes mellitus, however, evidence on the comparative effectiveness of these alternative drugs classes, particularly in relation to reducing HbA_{1c} levels, is insufficient to recommend a particular second line treatment.⁷ An international consensus statement⁹ and guidelines from the National Institute of Health and Care Excellence (NICE)⁷ therefore leaves the choice of second line treatment for most people with type 2 diabetes mellitus to clinicians and patients, which has led to wide variation across groups of primary care providers in England in the proportion of people prescribed each drug class.¹² Current NICE (2022) guidelines recommend other antidiabetic treatments, such as insulin based therapy and glucagon-like peptide-1 receptor agonists, only if HbA_{1c} levels are not controlled after second line treatment with oral antidiabetics.⁷ Hence in many countries, including England, the proportion of people with type 2 diabetes mellitus who are prescribed glucagon-like peptide-1 receptor agonists as second line treatment is low.^{10 12 13}

Most randomised controlled trials assessing the effectiveness and safety of SGLT-2 inhibitors and DPP-4 inhibitors have randomised groups to an active intervention or placebo comparator.¹⁴⁻²⁶ Therefore, although these trials reported fewer CVD and kidney events in people with and without type 2 diabetes mellitus allocated to SGLT-2 inhibitors, the results are difficult to apply to routine clinical practice, where the relevant populations and comparators differ.¹⁶⁻²⁴ Of the randomised controlled trials with an active comparator, some compared DPP-4 inhibitors with sulfonylureas²⁷⁻³⁰ or compared SGLT-2 inhibitors with sulfonylureas,³¹ but none compared all three drug classes. Thus the comparative effectiveness of SGLT-2 inhibitors versus alternative second line oral antidiabetic treatments on outcomes important to people with type 2 diabetes mellitus, particularly reduction in HbA_{1c} level, remains unclear. Results from previous observational studies comparing these treatments³²⁻³⁴ are at risk of bias from residual (unmeasured) confounding. Although a recent observational study³⁵ emulated some of the results of the GRADE (Glycemia Reduction Approaches In Diabetes: A Comparative Effectiveness

Study) randomised trial,^{29 36 37} neither the trial nor the observational study considered SGLT-2 inhibitors, which limits the applicability of the results to routine clinical practice.

Recent advances in real world data combined with developments in quantitative methods offer important opportunities for generating evidence on comparative effectiveness of treatments with direct relevance to clinical practice.³⁵ In this study, we illustrated the potential and challenges of using real world data from Clinical Practice Research Datalink (CPRD) for these purposes. We emulated the design of a hypothetical pragmatic randomised controlled trial by comparing three antidiabetic drug classes (sulfonylureas, DPP-4 inhibitors, and SGLT-2 inhibitors) of interest to the broad population of people with type 2 diabetes mellitus who, according to current NICE guidelines, are eligible for any of these second line treatments. We considered intermediate metabolic outcomes, particularly HbA_{1c} level, but also kidney and cardiovascular related complications. To reduce the risk of unmeasured confounding we used prescriber variation as an instrumental variable to estimate treatment effectiveness from routine data.^{38 39} Our study complements a recent target trial emulation that assessed the comparative effectiveness of alternative second line treatments using data from the Department of United States Veterans Affairs,⁴⁰ but which underrepresented female members of the population (<10%) and in the main analyses assumed that there was no unmeasured confounding.

We compared the effectiveness of the three most prescribed second line antidiabetic treatments in the UK according to metabolic and other clinical measures (changes from baseline in HbA_{1c} level, estimated glomerular filtration rate (eGFR), body mass index (BMI), and systolic blood pressure) and to adverse clinical endpoints (kidney and cardiovascular outcomes, and death).

Methods

Study design

We designed this study according to the target trial framework.⁴¹ Briefly, a target trial is a hypothetical randomised controlled trial for assessing comparative effectiveness from observational data that requires pre-specification of the main elements of a trial's protocol, including eligibility criteria, the respective treatment strategies, time zero, and an analysis plan.⁴¹ The target trial emulation reported in this paper is part of the PERMIT (PERsonalised Medicine for Intensification of Treatment) study, which prespecified the definition of the eligibility criteria and treatment strategies in the published versions of the study protocol⁴² and other elements of the target trial emulation in the statistical analysis plan.⁴³ Supplementary table 1 provides details to accompany this paper of how each of the standpoints were emulated (eligibility criteria, treatment assignment, initiation, and strategy, follow-up, outcomes, causal contrasts of interest, and analysis strategy).

We applied target trial principles to primary care data from CPRD to identify people with type 2 diabetes mellitus who had a similar prognosis before initiating any of the three second line antidiabetic treatments under comparison. CPRD covers about 20% of the UK population registered with general practices and includes longitudinal information on primary care diagnoses, prescriptions, personal information, and laboratory test results.^{44 45} Linkage from CPRD to Hospital Episode Statistics in-patient data was available for about 90% of participating practices in England. We accessed information from Hospital Episodes Statistics admitted patient care database on diagnoses, procedures, sociodemographic characteristics, and admission and discharge dates.⁴⁶ Rather than relying on a single data source to ascertain cardiovascular and kidney outcomes, we used linked data from CPRD-Hospital Episodes Statistics as these have been shown to improve capture of these events and reduce risks of misclassification.^{47 48} Information on each person's vital status was available through linkage to the Office for National Statistics (ONS) death records.^{49 50}

Study population

We defined the study population according to eligibility criteria, which had to be met before time zero (baseline) and was analogous to the time of randomisation in a randomised controlled trial. Time zero was defined by the date of the first prescription for any of the three oral second line treatments that were added to metformin (see supplementary table 1). We followed precedent research by including people with a diagnosis of type 2 diabetes mellitus who were aged 18 years or older,^{33 51} registered with a general practice in England, and who intensified treatment from first line to second line oral antidiabetic treatment between 1 January 2015 and 31 December 2020 with a first ever prescription of sulfonylureas, DPP-4 inhibitors, or SGLT-2 inhibitors added to metformin. Those eligible had to have at least one prescription for metformin monotherapy within 60 days before the first prescription for second line treatment, to ensure their use of metformin monotherapy was continuous before intensification. We excluded individuals with pregnancy recorded within 12 months before initiation of second line treatment and people whose last recorded eGFR was <30 mL/min/1.73m², since prescribing guidelines recommend different treatments for these groups. We also excluded people whose general practices had not consented to the required linkage of Hospital Episodes Statistics data. We followed precedent research in excluding those who were not prescribed metformin on the same day or within 60 days after initiating second line treatment,³³ as it is unlikely that their treatment with metformin continued. Supplementary tables 1 and 2 present detailed inclusion and exclusion criteria.

Treatments under comparison

We compared DPP-4 inhibitors with sulfonylureas and SGLT-2 inhibitors with sulfonylureas and DPP-4

inhibitors as second line oral antidiabetic treatments added to metformin. Information was extracted on the prescribed duration of each treatment and any subsequent antidiabetic treatment.

The study used an intention-to-treat approach so that individuals contributed to the treatment group to which they were assigned at baseline until the end of the follow-up period (see supplementary table 1), irrespective of the extent to which they adhered to the treatment prescribed. We defined the end of follow-up as the earliest of the date the general practice stopped contributing to CPRD, the date the individual left the general practice, the date of death, or the last date of available data (31 December 2021 for continuous outcomes or 31 March 2021 for time-to-event outcomes). We described the duration of second line and third line treatments by comparison group.

Covariates

We have previously described the covariates in detail,^{11 42} and these are summarised in supplementary table 3. Briefly, we defined patient sociodemographic characteristics (age, sex, ethnicity, index of multiple deprivation), time since diagnosis of type 2 diabetes mellitus, year of initiation of second line antidiabetic treatment, NHS region (East of England, London, Midlands, North East and Yorkshire, North West, South East, and South West),⁵² number of patients registered with the participants' general practice, smoking and alcohol intake status, relevant co-prescriptions (renin-angiotensin system inhibitors or statins) issued within 60 days before baseline, hospital admission (any) in the previous year, and comorbidities recorded at baseline (history of myocardial infarction, unstable angina, previous stroke, ischaemic heart disease, hypoglycaemia, heart failure, history of any cancer, history of proteinuria, advanced eye disease, lower limb amputation, and impaired kidney function (latest eGFR <60 mL/min/1.73m²). We also defined HbA_{1c}, systolic blood pressure, diastolic blood pressure, eGFR, and BMI⁵³ using the most recent measures recorded in primary care.

For the primary endpoint, change in HbA_{1c} level, we only considered the most recent measure within 180 days before time zero as the baseline measure in line with NICE guidance, which recommends that HbA_{1c} is measured every six months.⁷ For systolic and diastolic blood pressure and eGFR we followed previous research in considering the most recent measure within 540 days before baseline³³ (see supplementary table 3). We considered any values recorded in advance of these time windows as out-dated, and they were not used to define baseline characteristics. For BMI we followed a previously published algorithm in using the most recent measure available, which for most participants was within six months.⁵³

Outcomes

The primary outcome was the absolute change in HbA_{1c} (mmol/mol) level between baseline and one year after each prescription for second line treatment (HbA_{1c}

value at one year–HbA_{1c} value at baseline). Treatment groups were compared according to the mean change in HbA_{1c} level. We used the measurement closest in time to the one year follow-up time point and allowed for measures within ± 90 days, otherwise the measure was designated as missing.

Secondary outcomes included change in HbA_{1c} level at two years and change in BMI, systolic blood pressure, and eGFR at one year and two years.³³ We also reported the time to several first events before two years' follow-up: a $\geq 40\%$ decline in eGFR from baseline, which could be a marker for the rarer end stage kidney disease outcome⁵⁴; a major adverse kidney event, a composite outcome for the earliest of a decline in eGFR from baseline of 40%, end stage kidney disease, and all cause mortality⁵⁵; hospital admission for heart failure; major adverse cardiovascular event (MACE), a composite outcome for the earliest of myocardial infarction, stroke, or CVD death; and all cause mortality. We also reported time to myocardial infarction and stroke individually. Time to end stage kidney disease and CVD specific mortality could not be reported owing to the low number of events. Individuals were followed until they experienced the event of interest, died, or linked CPRD-Hospital Episodes Statistics data were no longer available (patient/general practice stopped contributing data to the CPRD or 31 March 2021). For these time-to-event measures, we only considered outcomes within the first two years in the base case, as it was anticipated that at later time points a high proportion of individuals would have censored or missing data. Supplementary table 4 provides details on all outcome definitions, including data sources.

Statistical analysis

We chose to use an instrumental variable analysis to help reduce the risk of confounding from unobserved baseline measures, such as diet and exercise before initiation of second line treatment (see supplementary methods, supplementary table 1, and supplementary figures 1A and 1B).³⁸ The instrumental variable was the primary care providers' tendency to prescribe the three classes of second line treatment. In England, most primary care clinicians work within a group, and over the study's timeframe this was defined as a clinical commissioning group (CCG), which informed health funding decisions for its respective geographical region. Some CCGs recommended that a relatively high proportion of people had second line treatment with sulfonylureas or DPP-4 inhibitors due in part to the higher cost of SGLT-2 inhibitors. We therefore defined CCGs rather than individual general practices as the unit for the instrumental variable, as this reflected decision making and was strongly associated with choice of second line treatment.^{11 12}

We also found wide variation across CCGs in the proportion of people prescribed each of the three classes of second line treatment (fig 1). This natural variation implied that people with a similar prognosis at baseline received a different second line treatment simply according to their CCG. We defined the tendency

to prescribe as the proportion of eligible people prescribed each second line treatment within the 12 months preceding the specific baseline (time zero) for each person. A valid instrument must meet four main conditions (see also the direct acyclic graph in supplementary figures 1A and 1B).³⁸ Firstly, the instrument must predict the treatment prescribed, which can be formally assessed.⁵⁶ Here, we assessed the relevance of the CCGs tendency to prescribe using a weak instrument test that is robust to heteroscedasticity and clustering by NHS region. Recent work has suggested that to meet the requirement that the instrument is of sufficient strength, the F statistic summarising the association between the instrumental variable and the treatment received must exceed 100.^{38 57} Secondly, the instrument must be independent of covariates that predict the outcomes of interest, which can be partially evaluated. We assessed the extent to which observed prognostic covariates differed across levels of the instrument (see supplementary figures 2A-2C). Thirdly, the instrument must have an effect on the outcomes only through the treatment received, which cannot be evaluated empirically. Large imbalances in measured covariates across levels of the tendency to prescribe would raise concerns about the second and third instrumental variable assumptions. We followed our prespecified protocol⁴² and the statistical analysis plan⁴³ and were guided by the direct acyclic graphs (see supplementary figures 1A and 1B) in choosing to adjust for measured contextual and temporal confounders in the second stage (outcome) regression. By including these contextual covariates in the second stage regression we were able to make weaker assumptions, that the tendency to prescribe was independent of the outcome and only had an effect on the outcome through the treatment received after adjusting for any differences in region, general practice size, and time period (see supplementary file). Fourthly, the instrumental variable analysis assumes monotonicity, which implies that as the levels of the instrumental variable change this should have the same direction of effect on the treatment prescribed across similar individuals. However, this assumption cannot be verified.⁵⁸ Indeed, in our study, we cannot observe the same treatment choice for a particular individual according to their attendance at two CCGs with different levels of prescribing preference for SGLT-2 inhibitors (versus DPP-4 inhibitors or sulfonylureas). For the population, this assumption implies that the average treatment choice must increase or decrease monotonically with the level of the instrumental variable.⁵⁹ Hence it is plausible to assume that if a group of patients whose CCG had a moderate preference for prescribing SGLT-2 inhibitors were prescribed this drug class, then a similar group of patients whose CCG had a stronger preference for prescribing SGLT-2 inhibitors would not be prescribed DPP-4 inhibitors or sulfonylureas.⁵⁹

We used the two stage residual inclusion method for the instrumental variable analysis,⁶⁰ which enabled us to assess comparative effectiveness across the full study populations of interest—that is, to report

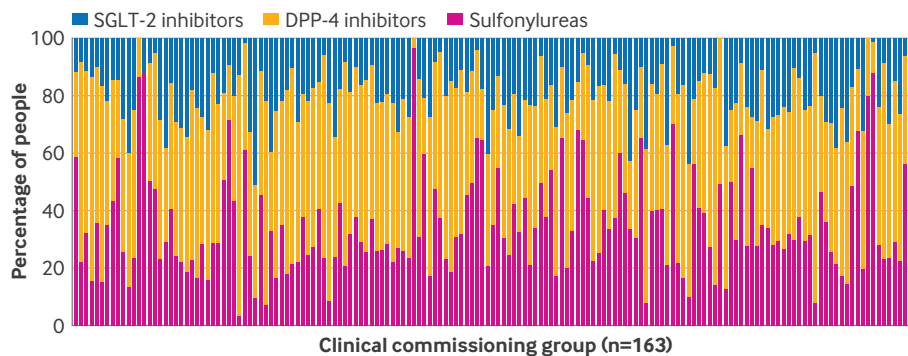


Fig 1 | Stacked bar chart illustrating variation in second line antidiabetic treatment prescribed among people included in the study at the clinical commissioning group level in England, 2014-20. DPP-4=dipeptidyl peptidase-4; SGLT-2=sodium-glucose cotransporter-2

average treatment effects while reducing the risk of bias from unmeasured confounding. The first stage models estimated the probabilities that each person was prescribed each treatment given their baseline covariates and their CCGs tendency to prescribe that treatment.⁶¹ The second stage outcome models then included generalised residuals from the first stage (propensity score) models. We estimated the outcome models by ordinary least squares for continuous outcomes (eg, HbA_{1c} level at one year) and by Cox proportional hazards models for time-to-event outcomes with an individual frailty.³² Models for both stages included all measured baseline covariates, with polynomials and covariate interactions selected through a post-double selection approach using least absolute shrinkage and selection operator regression⁶²⁻⁶⁴ (see supplementary methods table S1). The purpose of including person level covariates in the second stage (outcome regression) was to gain precision in estimating the relative treatment effects.

Some data were missing for outcomes (metabolic and other clinical measures) and baseline covariates (ethnicity, index of multiple deprivation, HbA_{1c}, systolic blood pressure, diastolic blood pressure, BMI, eGFR, smoking and alcohol intake status) because the participants' general practices either had not recorded these measures or had, but outside the requisite time window for a specific time point. At one year and two years, the percentages of missing values were, respectively, 33.7% and 36.4% for HbA_{1c}, 44.7% and 47.8% for BMI, 33.6% and 37.2% for systolic blood pressure, and 37.4% and 40.0% for eGFR. For some people, a measurement that was not available at a particular time point (eg, two years) was available at other time points (eg, one year and three years) (see supplementary methods table S2). It was also possible that at any time point, one measure (eg, BMI) was not available, whereas other measures (eg, HbA_{1c}, systolic blood pressure, and eGFR) were available.

We chose to handle all missing baseline and longitudinal outcome data by multiple imputation⁶⁵ with chained equations.⁶⁶ This approach assumed

data were missing at random. The imputation of each longitudinal outcome at a given time point used all relevant information, including measurements of the same outcome at other time points. This use of auxiliary information can help the study recover more accurate estimates of the unknown outcome values.⁶⁷ This also ensured our study population was comparable at each time point. Partially observed covariates and outcomes^{67 68} were multiple imputed by predictive mean matching with 10 donors,⁶⁹ producing five imputed datasets. The number of imputations was driven by the need to balance computational time with improved inference from increasing the number of imputations (see supplementary methods for further details). The imputation models developed for each covariate were congenial with the form of outcome⁷⁰ (continuous or time to event). For the time-to-event endpoints, it was assumed no data were missing. All imputation models were stratified by second line treatment (DPP-4 inhibitors, SGLT-2 inhibitors, sulfonylureas) and by whether the individual died or was censored before the relevant study end date (see supplementary methods).

We reported differences between the comparison groups according to absolute change in outcomes between baseline and follow-up for continuous measures, and according to time-to-event measures. We reported results overall and according to whether patients had or did not have CVD (at least one of previous myocardial infarction, previous stroke, heart failure, ischaemic heart disease, or unstable angina) recorded before initiation of second line treatment. To recognise statistical uncertainty in the estimates of treatment effects, the data were bootstrapped 500 times, stratified by CCG, treatment group, and death and censoring status to maintain the structure of the original sample across replicates. Within each bootstrap resample we implemented multiple imputation with chained equations,^{71 72} with Rubin's first rule⁶⁵ applied across the five imputed datasets to obtain overall treatment effects for each bootstrap sample, which we then used to estimate variances and

calculate *t* based bootstrap confidence intervals (CI). The imputation procedure and time-to-event analyses were performed with multiple imputation with chained equations and the survival package^{73 74} in R 4.2.2, respectively,⁷⁵ and the analysis of the clinical measures in Stata 17.⁷⁶

Alternative analyses

We undertook alternative analyses to check the impact of different statistical assumptions on our results. Firstly, we applied complete case analysis rather than multiple imputation with chained equations (base case) to examine whether the results were robust when alternative approaches were applied to handle missing data. Secondly, we applied two stage least squares (continuous outcomes), multivariable linear regression (continuous outcomes), and Cox regression analysis (time to event), adjusting for all measured baseline covariates, to assess the sensitivity of our approach to confounding adjustment. Thirdly, we extended the follow-up period to five years rather than two years. Fourthly, in additional analyses that were not prespecified, we further checked the impact of applying approaches that, as with multivariable regression, assumed no unmeasured confounding but can be less sensitive to the form of outcome regression model. We applied two approaches based on propensity scores—inverse probability of treatment weighting⁷⁷ and inverse probability of treatment weighting with regression adjustment (weighted regression hereafter),⁷⁸ with non-stabilised and stabilised weights.⁷⁹ We also used asymmetrical trimming to understand any effects of large weights in the weighted regression analysis.^{80 81} The weighted regression has the so called double robustness property, in that, subject to the assumption of no unobserved confounding, it can still provide consistent estimates provided either the propensity score or the regression model is correctly specified.^{78 82} The multivariable regression analyses, the inverse probability of treatment weighting, and the weighted regression analyses all estimate the average treatment effects as in the base case. We undertook the alternative analyses on the complete cases only.

Patient and public involvement

Patient and public involvement advisors, including a coauthor on this paper (PC), helped inform the design and proposed analysis, including the choice of outcome measures. We will reconvene a patient and public involvement workshop to discuss the study findings and co-produce a lay summary that will be available on the PERMIT study website.⁸³

Results

Study population and baseline characteristics

The study population included 75 739 people with type 2 diabetes mellitus who initiated second line oral antidiabetic treatment with sulfonylureas, DPP-4 inhibitors, or SGLT-2 inhibitors and met all eligibility criteria (fig 2). Of these, 25 693 (33.9%) initiated treatment with sulfonylureas, 34 464 (45.5%) with

DPP-4 inhibitors, and 15 582 (20.6%) with SGLT-2 inhibitors, in addition to metformin. Supplementary table 5 reports the frequencies of prescribing for each drug within each drug class. The drugs most commonly prescribed within each drug class were gliclazide (sulfonylurea), sitagliptin (DPP-4 inhibitor), and empagliflozin (SGLT-2 inhibitor). People prescribed SGLT-2 inhibitors were younger (56 (standard deviation (SD) 11) years) than those prescribed DPP-4 inhibitors (62 (SD 12) years) or sulfonylureas (60 (SD 13) years) (table 1). The baseline mean HbA_{1c} level was higher for people prescribed sulfonylureas (81 (SD 22) mmol/mol) compared with those prescribed DPP-4 inhibitors (72 (SD 16) mmol/mol) or SGLT-2 inhibitors (75 (SD 17) mmol/mol), and a lower proportion of people prescribed SGLT-2 inhibitors had comorbidities—for example, 17.2% (n=2680) of those prescribed SGLT-2 inhibitors had prevalent CVD compared with 22.8% (n=5858) of those prescribed sulfonylureas and 23.5% (n=8108) prescribed DPP-4 inhibitors. The proportion of people prescribed SGLT-2 inhibitors increased from 7.3% in 2015 to 24.9% in 2020. The median time between recorded BMI and the index date was 19 days (interquartile range (IQR) 0-140 days).

Within two years of follow-up, the median (IQR) time prescribed second line antidiabetic treatment was lower for those using sulfonylureas (248 (IQR 67-671) days) compared with DPP-4 inhibitors (345 (IQR 96-730) days) and SGLT-2 inhibitors (328 (IQR 84-730) days). The proportion of participants who switched to a third line treatment within two years of the index date was 58.8% (sulfonylureas, n=15 107), 51.5% (DPP-4 inhibitors, n=17 749), and 52.5% (SGLT-2 inhibitors, n=8184), with metformin monotherapy the most common third line treatment for all three comparison groups (see supplementary table 6). In each comparison group, the proportions of people whose third line treatment was triple therapy were 25.1% (sulfonylureas), 31.7% (DPP-4 inhibitors), and 21.8% (SGLT-2 inhibitors).

Empirical assessment of instrumental variable assumptions

The tendency to prescribe met a major requirement for being a valid instrumental variable, in that it was strongly associated with the second line treatment prescribed (assumption 1), with accompanying F statistics of 1902 for DPP-4 inhibitors and 1935 for SGLT-2 inhibitors, which indicated that the instrumental variable was of sufficient strength ($F > 100$).^{38 57} The measured potential confounders were balanced across levels of the tendency to prescribe (assumption 2), aside from time period, which was included within the covariate adjustment of the instrumental variable analysis (see supplementary figures 2A-2C).

Intermediate metabolic and other clinical measures

The crude change in mean HbA_{1c} level from baseline to one year follow-up among people with observed follow-up measures was greatest for those prescribed sulfonylureas (−18 mmol/mol) compared with DPP-

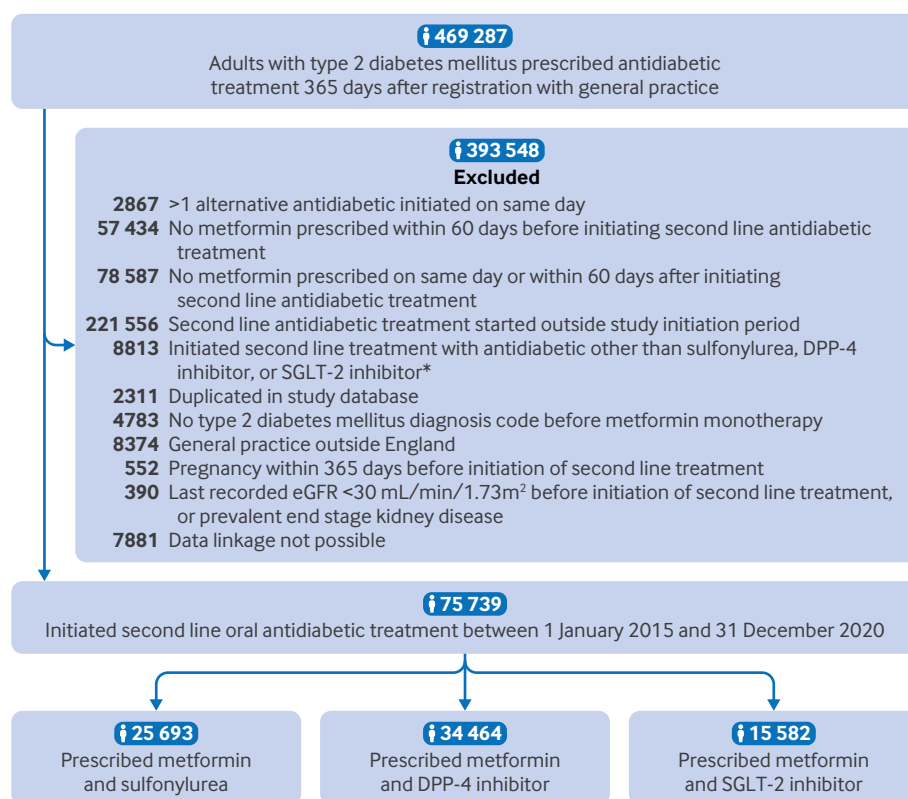


Fig 2 | Flow of study population. *Other antidiabetic drugs prescribed as second line treatment included thiazolidinediones, glucagon-like peptide 1 receptor agonists, and insulin. DPP-4=dipeptidyl peptidase-4; SGLT-2=sodium-glucose cotransporter-2

4 inhibitors (–10 mmol/mol) and SGLT-2 inhibitors (–14 mmol/mol; fig 3, also see supplementary figure 3). Of those people not censored by one year follow-up (n=72 066), 33.7% were missing HbA_{1c} values at this time point (see supplementary methods table 2). Although levels of missing data were higher for those time points that occurred after the onset of the covid-19 pandemic, the levels of missing data remained similar across the comparison groups (see supplementary table 7).

The crude changes in mean BMI and systolic blood pressure from baseline were small across all time points (fig 3, also see supplementary figure 3). The crude change in mean eGFR from baseline to one year follow-up was similar across the three second line treatments of interest (–2 mL/min/1.73m²), with smaller decreases in mean eGFR across subsequent follow-up periods among people prescribed SGLT-2 inhibitors rather than sulfonylureas or

Table 1 | Baseline characteristics of the primary-secondary care linked study population, stratified by prescribed second line antidiabetic treatment. Values are number (percentage) unless specified otherwise

Characteristics	Sulfonylureas (n=25 693)	DPP-4 inhibitors (n=34 464)	SGLT-2 inhibitors (n=15 582)
Female participants	9840 (38.3)	13 456 (39.0)	5996 (38.5)
Mean (SD) age (years)	60 (13)	62 (12)	56 (11)
Ethnicity:			
White	19 665 (76.5)	27 308 (79.2)	12 543 (80.5)
South Asian	3522 (13.7)	4616 (13.4)	1961 (12.6)
Black	1625 (6.3)	1451 (4.2)	542 (3.5)
Mixed/Other	534 (2.1)	612 (1.8)	231 (1.5)
Missing	347 (1.4)	477 (1.4)	305 (2.0)
Index of multiple deprivation fifth:			
1 (least deprived)	3619 (14.1)	5161 (15.0)	2604 (16.7)
2	4504 (17.5)	6175 (17.9)	2793 (17.9)
3	4955 (19.3)	6642 (19.3)	2953 (19.0)
4	6152 (23.9)	7677 (22.3)	3408 (21.9)
5 (most deprived)	6449 (25.1)	8785 (25.5)	3815 (24.5)
Missing	14 (0.1)	24 (0.1)	9 (0.1)

(Continued)

Table 1 | Continued

Characteristics	Sulfonylureas (n=25 693)	DPP-4 inhibitors (n=34 464)	SGLT-2 inhibitors (n=15 582)
Year second line antidiabetic treatment was initiated:			
2015	6996 (27.2)	4958 (14.4)	1145 (7.3)
2016	5221 (20.3)	6057 (17.6)	1525 (9.8)
2017	4260 (16.6)	6309 (18.3)	2178 (14.0)
2018	3562 (13.9)	6771 (19.6)	2912 (18.7)
2019	3112 (12.1)	5995 (17.4)	3939 (25.3)
2020	2542 (9.9)	4374 (12.7)	3883 (24.9)
Median (IQR) No of years receiving metformin monotherapy	2.9 (1.1-5.4)	3.6 (1.7-6.3)	2.8 (1.3-5.2)
Median (IQR) mean No of patients registered with general practice	9690 (6250-13 628)	9971 (6538-13 795)	10 143 (6896-13 881)
Last HbA _{1c} value recorded before index date:			
Mean (SD) HbA _{1c} (mmol/mol)	81 (22)	72 (16)	75 (17)
Mean (SD) HbA _{1c} (%)	9.1 (2.1)	8.2 (1.5)	8.5 (1.6)
HbA _{1c} level (mmol/mol):			
<53	713 (2.8)	1053 (3.1)	515 (3.3)
53-74	10 818 (42.1)	21 870 (63.5)	8410 (54.0)
≥75	12 579 (49.0)	10 398 (30.2)	6134 (39.4)
Missing	1583 (6.2)	1143 (3.3)	523 (3.4)
Last blood pressure measure recorded before index date:			
Mean (SD) systolic blood pressure (mm Hg)	132 (14)	132 (14)	133 (14)
Mean (SD) diastolic blood pressure (mm Hg)	78 (9)	77 (9)	80 (9)
Hypertensive status based on last recorded blood pressure measure:			
Normotensive	7123 (27.7)	9424 (27.3)	3664 (23.5)
Hypertensive	18 525 (72.1)	25 002 (72.5)	11 906 (76.4)
Missing	45 (0.2)	38 (0.1)	12 (0.1)
Mean (SD) BMI at index date	31.5 (6.6)	32.2 (6.5)	35.1 (7.0)
BMI at index date:			
Under/normal weight	2718 (10.6)	2782 (8.1)	394 (2.5)
Overweight	8110 (31.6)	10 180 (29.5)	2867 (18.4)
Obese	14 702 (57.2)	21 375 (62.0)	12 283 (78.8)
Missing	163 (0.6)	127 (0.4)	38 (0.2)
Mean (SD) last recorded eGFR before index date (mL/min/1.73m ²)	91 (19)	88 (19)	97 (15)
eGFR category (mL/min/1.73m ²):			
Stage 1-2 (≥60)	23 282 (90.6)	30 823 (89.4)	15 186 (97.5)
Stage 3a-3b (30-59)	1770 (6.9)	3199 (9.3)	161 (1.0)
Missing	641 (2.5)	442 (1.3)	235 (1.5)
Comorbidities:			
Prevalent CVD	5858 (22.8)	8108 (23.5)	2680 (17.2)
Lower limb amputation	227 (0.9)	265 (0.8)	76 (0.5)
Heart failure	1457 (5.7)	2007 (5.8)	598 (3.8)
History of myocardial infarction	1644 (6.4)	2226 (6.5)	842 (5.4)
Previous stroke	1378 (5.4)	1678 (4.9)	512 (3.3)
Ischaemic heart disease	4572 (17.8)	6538 (19.0)	2175 (14.0)
Unstable angina	777 (3.0)	1099 (3.2)	362 (2.3)
History of any cancer	4254 (16.6)	5397 (15.7)	1447 (9.3)
Blindness	425 (1.7)	527 (1.5)	140 (0.9)
History of hypoglycaemia	260 (1.0)	302 (0.9)	129 (0.8)
Proteinuria	3658 (14.2)	4679 (13.6)	1585 (10.2)
Co-prescriptions:			
RAS inhibitor	12 584 (49.0)	18 911 (54.9)	8108 (52.0)
Statin	17 729 (69.0)	25 690 (74.5)	10 838 (69.6)
Smoking status:			
Non-smoker	5720 (22.3)	7455 (21.6)	3562 (22.9)
Former smoker	12 640 (49.2)	18 009 (52.3)	7865 (50.5)
Current smoker	7327 (28.5)	8992 (26.1)	4154 (26.7)
Missing	6 (0.0)	8 (0.0)	<5
Alcohol intake:			
Non-drinker	3192 (12.4)	3716 (10.8)	1630 (10.5)
Former drinker	7248 (28.2)	10 009 (29.0)	4179 (26.8)
Current drinker	14 899 (58.0)	20 367 (59.1)	9582 (61.5)
Missing	354 (1.4)	372 (1.1)	191 (1.2)

BMI=body mass index; CVD=cardiovascular disease; DPP-4=dipeptidyl peptidase-4; eGFR=estimated glomerular filtration rate; IQR=interquartile range; RAS=renin-angiotensin system; SD=standard deviation; SGLT-2=sodium-glucose cotransporter-2.

DPP-4 inhibitors (fig 3, also see supplementary figure 3).

Figure 4 presents the results from the instrumental variable analysis, which reduces the risk of

confounding, and after applying multiple imputation with chained equations to handle the missing data. The results apply to the full study population. Strong evidence was found for SGLT-2 inhibitors being more

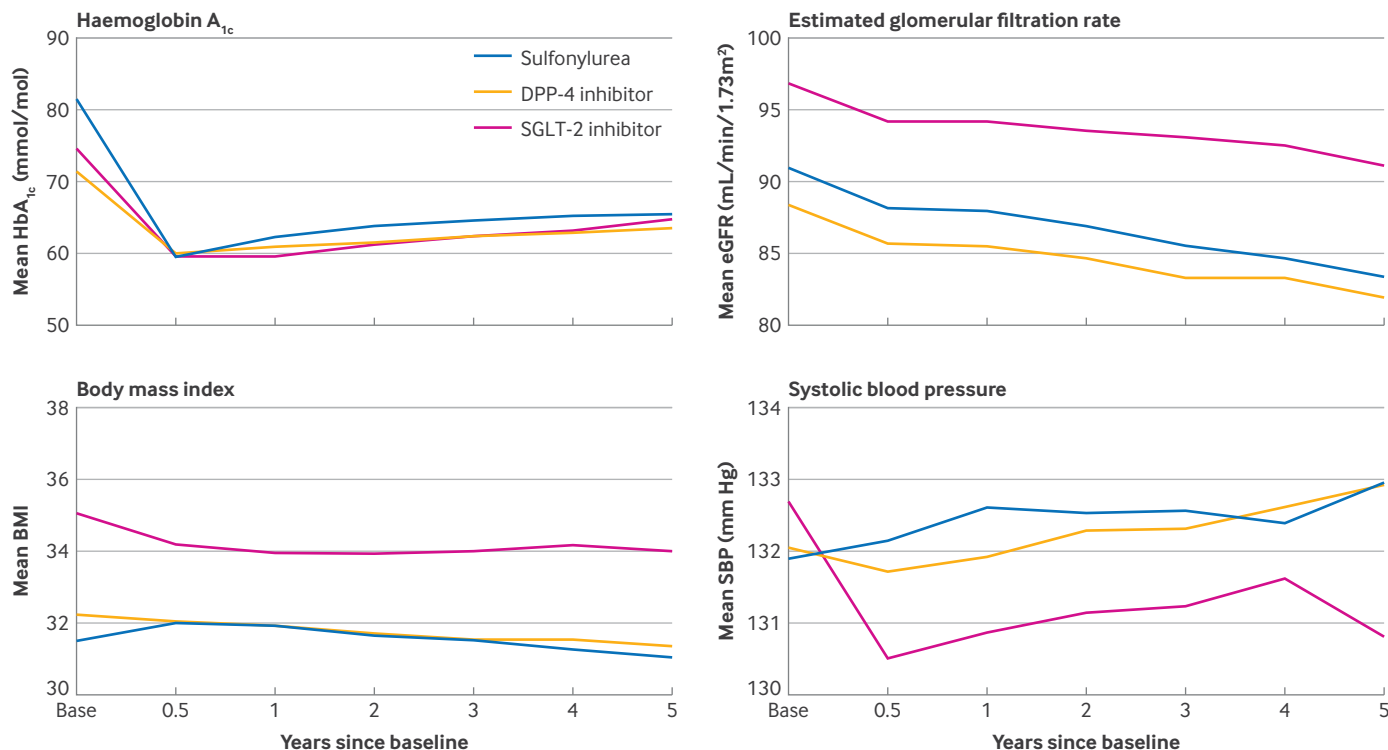


Fig 3 | Mean HbA_{1c} (mmol/mol), estimated glomerular filtration rate (mL/min/1.73m²), body mass index, and systolic blood pressure (mm Hg) at each follow-up time point of interest, stratified by treatment group. DPP-4=dipeptidyl peptidase-4; eGFR=estimated glomerular filtration rate; HbA_{1c}=glycated haemoglobin A_{1c}; SBP=systolic blood pressure; SGLT-2=sodium-glucose cotransporter-2

effective in reducing HbA_{1c} levels between baseline and one year follow-up, with a mean reduction of -2.5 mmol/mol (95% CI -3.7 to -1.3) compared with sulfonylureas and -3.2 mmol/mol (-4.6 to -1.8) compared with DPP-4 inhibitors (fig 4, also see supplementary table 8). After accounting for confounding and missing data, SGLT-2 inhibitors were more effective in improving BMI and systolic blood pressure (fig 4). People prescribed SGLT-2 inhibitors showed a greater reduction in BMI between baseline and one year, with a mean difference of -1.6 (95% CI -1.7 to -1.4) compared with sulfonylureas and -0.8 (-1.0 to -0.7) compared with DPP-4 inhibitors. For systolic blood pressure, the mean difference was -2.1 mm Hg (95% CI -3.1 to -1.0) compared with sulfonylureas and -1.8 mm Hg (-3.0 to -0.5) compared with DPP-4 inhibitors, with these improvements maintained at two years follow-up. SGLT-2 inhibitors led to a slower decline in eGFR at two years follow-up compared with sulfonylureas (mean difference 1.4 mL/min/1.73m², 95% CI 0.5 to 2.3), but not compared with DPP-4 inhibitors (0.0 mL/min/1.73m², -1.1 to 1.0).

Kidney, cardiovascular, and mortality outcomes

People prescribed SGLT-2 inhibitors had lower crude rates of all adverse kidney, cardiovascular, and mortality events compared with those prescribed sulfonylureas and DPP-4 inhibitors (see supplementary

table 9 and supplementary figures 4-9). After reducing the risk of confounding and addressing the missing data, we found that over two years follow-up (base case), SGLT-2 inhibitors were more effective in preventing a $\geq 40\%$ decline in eGFR from baseline versus sulfonylureas (hazard ratio 0.42 , 95% CI 0.22 to 0.81), but the estimated hazard ratios for SGLT-2 inhibitors compared with DPP-4 inhibitors were highly uncertain (0.64 , 0.29 to 1.43) (fig 5). The rates of admission to hospital for heart failure were lower for SGLT-2 inhibitors compared with sulfonylureas (0.46 , 0.20 to 1.05) and with DPP-4 inhibitors (0.32 , 0.12 to 0.85). For the other endpoints, we found no evidence of a difference in the comparative effectiveness of the second line antidiabetic treatments (fig 5, also see supplementary table 10). We found no evidence that having CVD before starting second line treatment was associated with modified relative effectiveness of these three treatments (see supplementary tables 11 and 12).

Alternative analyses

The findings of the complete case analyses were similar when applying multiple imputation to deal with missing data (see supplementary tables 10 and 13). The results were also similar if the risk of confounding was dealt with using two stage least squares, an alternative instrumental variable approach (see supplementary table 14). The regression analyses that assumed

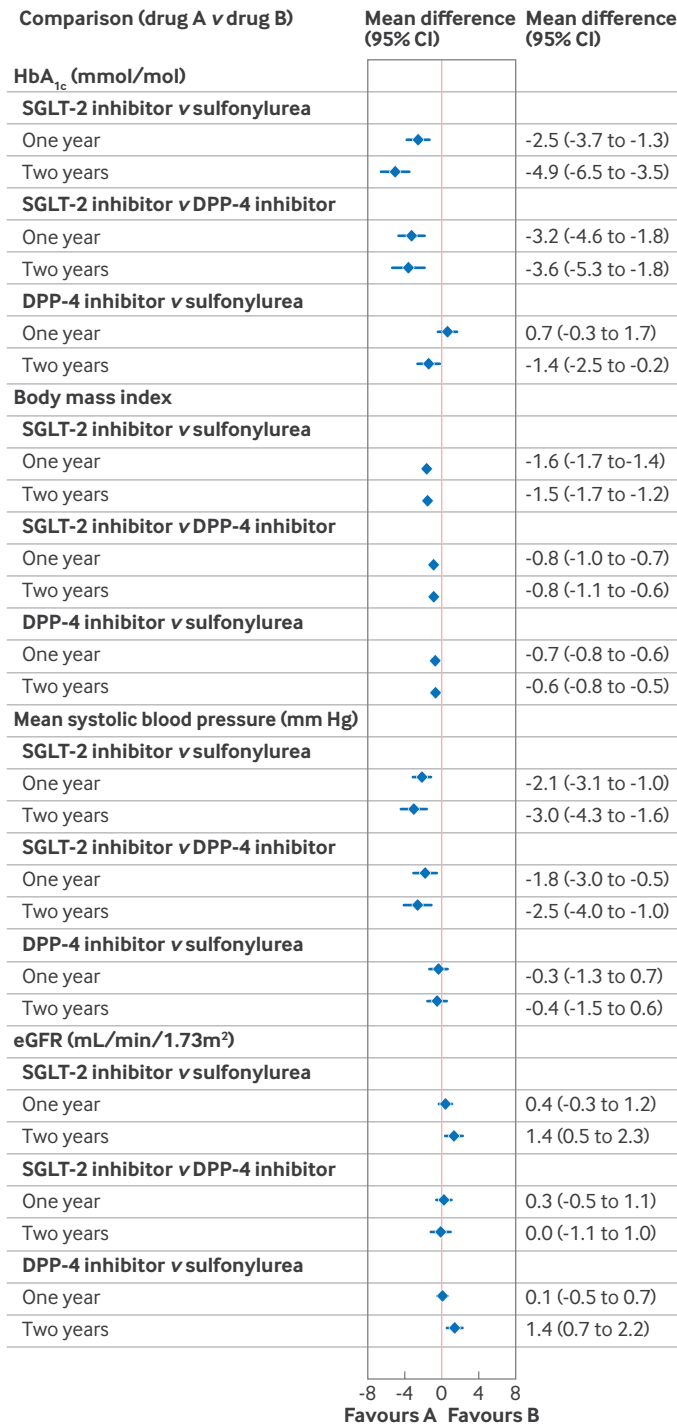


Fig 4 | Forest plot showing differences in the mean change between baseline and one year or two years follow-up in continuous clinical measures comparing second line antidiabetic treatments from the instrumental variable analysis to reduce the risk of confounding, and with multiple imputation to account for missing data, which assumes data are missing at random. CI=confidence interval; DPP-4=dipeptidyl peptidase-4; eGFR=estimated glomerular filtration rate; HbA_{1c}=glycated haemoglobin A_{1c}; SGLT-2=sodium-glucose cotransporter 2

no unmeasured confounders existed reported that the benefits of SGLT-2 inhibitors were greater than for the base case and more precisely estimated (see supplementary tables 10 and 15). When the study time frame was extended to five years, the gains after

initial receipt of SGLT-2 inhibitors were maintained, although by this time point few people had complete follow-up information or were still prescribed the same second line treatment (see supplementary tables 6, 8, 10, and 13-15). The results were similar to the main analyses if inverse probability of treatment weighting or weighted regression were used to reduce the risk of confounding due to observed baseline covariates (see supplementary tables 16-20 and supplementary figures 10 and 11).

Discussion

In this comparative effectiveness study, we found that second line treatment with SGLT-2 inhibitors for people with type 2 diabetes mellitus was more effective than sulfonylureas or DPP-4 inhibitors in reducing mean HbA_{1c} levels, BMI, and systolic blood pressure after the risk of confounding was reduced using an instrumental variable analysis. SGLT-2 inhibitors were also more effective at reducing the hazards of hospital admission for heart failure (compared with DPP-4 inhibitors) and ≥40% decline in eGFR (compared with sulfonylureas). We did not find strong evidence for other meaningful differences for the other study endpoints over the two year study period.

A major concern of any study aiming to assess comparative effectiveness from routine data is bias from confounding, particularly unmeasured prognostic differences between comparison groups. This risk of bias can never be eliminated. A crucial advantage of our study design, however, was that it followed recommended methods of target trial emulation in prespecifying the population eligibility criteria, time zero, treatment comparisons, outcomes, and analyses.^{41 84 85} In our main analysis we used an instrumental variable to further reduce the risk of residual confounding. We were therefore able to provide useful evidence about the comparative effectiveness of these three treatments as they were prescribed in routine clinical practice for a diverse population of people with type 2 diabetes mellitus.

The aim of the PERMIT study was to assess the relative effectiveness of the three most common second line treatments for an unselected population in routine clinical practice. In contrast, published randomised controlled trials have aimed to show the safety and efficacy of one of these drug classes compared with placebo in selected populations. For the comparison of SGLT-2 inhibitors versus placebo, published randomised controlled trials do not include general populations of people with type 2 diabetes mellitus who meet the eligibility criteria of national guidelines for these three second line treatments (see supplementary table 21).⁷ It is therefore challenging to compare the results of the PERMIT study with those of the published randomised controlled trials.

In supplementary tables 21-22, we describe the results of the PERMIT study alongside those of the corresponding randomised controlled trials for common endpoints such as hospital admission for heart failure, MACE, major adverse kidney events, and

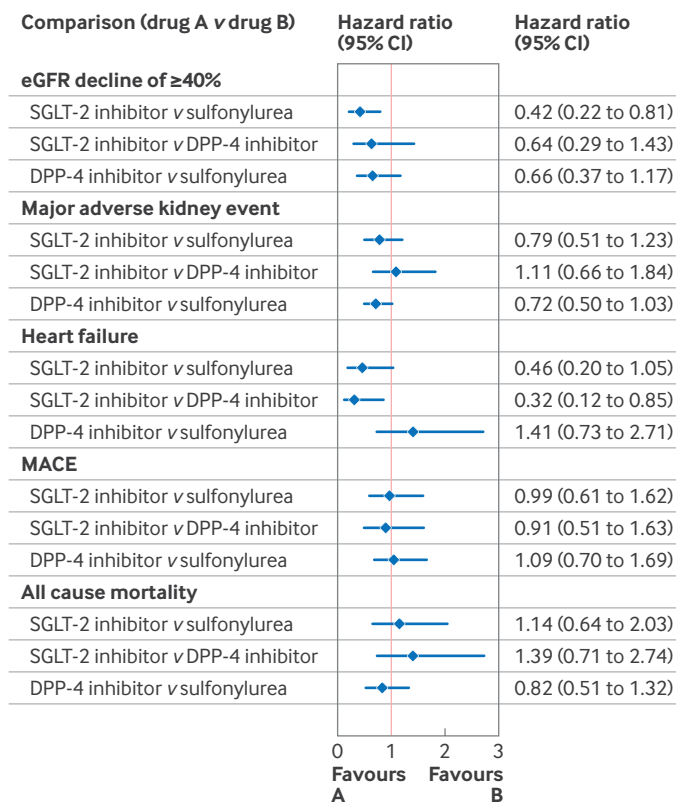


Fig 5 | Forest plot showing adjusted hazard ratios of cardiovascular disease and kidney outcomes comparing second line antidiabetic treatments from the instrumental variable analysis to reduce the risk of confounding, and with multiple imputation to account for missing data, which assumes data are missing at random. CI=confidence interval; MACE=major adverse cardiovascular event (composite for the earliest of myocardial infarction, stroke, or cardiovascular death); eGFR=estimated glomerular filtration rate; DPP-4=dipeptidyl peptidase-4; SGLT-2=sodium-glucose cotransporter 2

all cause death. We found that the point estimates for the PERMIT target trial emulation fall within the estimated 95% CI of the corresponding treatment effects reported in the randomised controlled trials—that is, they met previously defined criteria for agreement⁸⁵ (see supplementary table 22). This concordance also applied to the few published randomised controlled trials, including the GRADE trial,^{29,36} that compared two active treatments—DPP-4 inhibitors and sulfonylureas for general populations of people with type 2 diabetes mellitus. Unlike the GRADE trial, the PERMIT study did not exclude people with HbA_{1c} levels outside the range 6.8–8.5%. A previous target trial emulated the GRADE trial in applying strict eligibility criteria, but, unlike our study, it was unable to investigate MACE, heart failure, and all cause mortality owing to low event rates from a small study population. Our larger study found protective effects of SGLT-2 inhibitors for heart failure compared with DPP-4 inhibitors, similar to meta-analyses of randomised controlled trials^{86,87} and observational studies.³² However, even with this relatively large sample, the number of people followed over the full follow-up period was insufficient to detect other clinically important differences for outcomes

such as major adverse kidney events, and to investigate end stage kidney disease and CVD specific mortality individually.

In our alternative analysis, we made the common assumption of no unmeasured confounding, and found that after adjusting for all measured confounders, SGLT-2 inhibitors were associated with greater improvement in all endpoints, including all cause mortality. People prescribed SGLT-2 inhibitors, however, had fewer comorbidities and were likely to be healthier according to unmeasured baseline characteristics. A previous study that considered uptake of SGLT-2 inhibitors as second line antidiabetic treatment also reported that compared with people who received sulfonylureas or DPP-4 inhibitors, those who received SGLT-2 inhibitors were healthier and at lower risk of all cause death.³⁴ For an endpoint such as all cause death, it is particularly challenging to capture all the potential confounders from routine data sources (see supplementary figure 1B). For this endpoint, important potential confounders include the individual's overall health, diet, exercise, and lifestyle before second line treatment. If an instrumental variable is valid, it reduces the risk of bias from these unmeasured confounders, whereas approaches such as regression do not. Hence, the finding from the regression analysis that within the two year follow-up period SGLT-2 inhibitors were associated with reduced hazards of all cause mortality compared with sulfonylureas or DPP-4 inhibitors could reflect these unmeasured baseline differences (ie, residual confounding).

Strengths and weaknesses and comparison with other studies

In this study, we directly compared the three most commonly prescribed second line antidiabetic drug treatments using a large, linked dataset that is representative of the UK primary care population in terms of age and sex.^{44,45} Our direct comparison of sulfonylureas, DPP-4 inhibitors, and SGLT-2 inhibitors contrasts with previous trials^{16–18,20,26,29,36} and meta-analyses,^{86,87} which did not include an active second line treatment as a comparator. We did not restrict the study population to those with baseline HbA_{1c} levels in a particular range, as has been done in many previous randomised controlled trials.^{16–19} This study therefore includes people with a broader range of glycaemic control, which is reflective of the UK primary care population with type 2 diabetes mellitus.

We add to the evidence reported in previous observational studies,^{33–35,40,88} which make direct comparisons between antidiabetic treatments, by using an instrumental variable analysis as the main analysis to reduce the risk of confounding from both measured and unmeasured baseline confounders, and we provide evidence on comparative effectiveness for those three drug classes that are most commonly prescribed in publicly funded health systems for a general population of people with type 2 diabetes mellitus. We investigated not only intermediate metabolic and other clinical measures but also adverse

kidney and cardiovascular events, which are important to patients. The benefits we observed of SGLT-2 inhibitors improving HbA_{1c} levels, BMI, and systolic blood pressure and reducing the risks of admission to hospital for heart failure (compared with DPP-4 inhibitors) and $\geq 40\%$ decline in eGFR (compared with sulfonylureas) are indicative of a causal mechanism that has some biological plausibility.

Our directed acyclic graphs provided a framework for the analysis, which recognised that second line treatment with SGLT-2 inhibitors in routine practice could improve any of the intermediate clinical endpoints listed and which may in turn lead to reduced risks of subsequent events. In particular, the pharmacological action of SGLT-2 inhibitors—namely, reducing blood pressure and cardiac preload and after load through diuretic mechanisms,⁸⁹ would imply protective effects on hospital admissions for heart failure and on kidney endpoints; however, this would not necessarily translate to immediate protective effects during an ST elevated myocardial infarction or acute rupture of a coronary plaque.

We acknowledge limitations in our study. We did not consider glucagon-like peptide-1 receptor agonists because this class was rarely prescribed as a second line antidiabetic treatment in the UK during the study period,^{12 13} and they are still not recommended as second line treatment for people with type 2 diabetes mellitus.¹² The prescribing of glucagon-like peptide-1 receptor agonists is increasing in the US, however, and warrants further study as the number of people prescribed these drugs increases in routinely collected data. Our instrumental variable analysis relied on three major assumptions. Although we were able to empirically show that the instrument strongly predicts receipt of treatment (assumption 1), we could only partially evaluate whether the instrument was balanced across confounders (assumption 2). We adjusted for measured confounders in the second stage of the regression model to account for any residual imbalances across levels of the instrument, in particular according to time period and contextual measures such as region and general practice size. If, however, assumption 2 is not met, then unmeasured confounders would be imbalanced across levels of the instrument, leading to biased estimates. We must also assume that the instrument, the tendency to prescribe, does not directly impact outcomes except through the treatment prescribed (assumption 3). We could not test this assumption, and it is possible it would be violated if, for example, after adjusting for region and practice size, those CCGs that had a higher tendency to prescribe SGLT-2 inhibitors also delivered higher quality of care that improved outcomes independent of the second line treatment prescribed.

The PERMIT study used routine data, and the requisite outcome data were not available for all those included. For continuous measures, the proportion of people with missing values at the one year time point ranged from 33.7% (HbA_{1c}) to 44.7% (BMI). In the main analysis we dealt with these missing data for

all the continuous outcomes along with any missing information on covariates with multiple imputation, and we undertook complete case analysis as alternative analyses. The results from these alternative approaches that make different underlying assumptions about why the data were missing were similar. For the time-to-event endpoints, we used linked primary and secondary care and ONS death datasets to ascertain cardiovascular, kidney, and mortality outcomes to improve the capture of events, rather than relying on a single source. However, a limitation shared with other target trial emulations using routine data is that it is not known if data on events pertaining to kidney disease or CVD is missing. People may experience an event that is diagnosed and recorded in outpatient clinics but not recorded in the linked primary-secondary care (inpatient) data. For major events such as myocardial infarction or stroke, levels of under-recording in the linked data are likely to be small and similar across the comparison groups and lead to reduced statistical power rather than bias in the estimates of relative effectiveness.

Although the study did consider endpoints up to five years after initiation of second line treatment, by this time point levels of missing data were high (from 46.9% for HbA_{1c} to 59.4% for BMI), and after two years most people will have stopped their second line treatment. Hence, although we have reported results for the prespecified five year endpoint, caution is needed when interpreting these results, given the levels of missing data.

Policy implications

This study provides evidence that SGLT-2 inhibitors might offer clinically important benefits when provided in routine clinical practice compared with common alternative oral antidiabetic drugs that are added to metformin for people with type 2 diabetes mellitus. These findings apply to a wide range of people with type 2 diabetes mellitus and therefore complement the evidence available from randomised controlled trials¹⁶⁻²⁴ and previous studies that have emulated trials.^{35 40} In recent updated guidelines, NICE and other health technology assessment agencies have published guidance and guidelines that are neutral about the use of SGLT-2 inhibitors versus DPP-4 inhibitors versus sulfonylureas as second line treatments, except for people at high risk of CVD, or for people with pre-existing CVD, including heart failure, or with kidney disease. For these subgroups, SGLT-2 inhibitors are recommended in addition to metformin. Our study reported similar advantages for SGLT-2 inhibitors (compared with sulfonylureas and DPP-4 inhibitors) as second line treatments for people who did not have pre-existing CVD as well as for those who did have CVD before second line treatment. Future guidelines could draw from this study and related evidence to also recommend SGLT-2 inhibitors for those without CVD, including those at relatively low risk of subsequent CVD.

More work is needed to understand the long term effectiveness and cost effectiveness of increasing the

use of SGLT-2 inhibitors for people with type 2 diabetes mellitus. Future research can use the information from this study to predict whether SGLT-2 inhibitors can lead to sufficient improvement in long term outcomes—for example, from reduced incidence and costs of complications such as retinopathy, amputation, or end stage kidney disease, to justify any additional costs. Further research is also required to assess the comparative effectiveness of glucagon-like peptide-1 receptor agonists with the three alternative second line oral antidiabetic treatments among people with type 2 diabetes mellitus, and to assess how best to personalise the order in which these treatments are prescribed.

Conclusions

We found that for a broad population of people with type 2 diabetes mellitus, SGLT-2 inhibitors were more effective second line treatments in routine clinical practice compared with DPP-4 inhibitors or sulfonylureas in improving HbA_{1c} levels, BMI, and systolic blood pressure. SGLT-2 inhibitors were also found to be more effective at reducing the hazards of hospital admission for heart failure (compared with DPP-4 inhibitors) and $\geq 40\%$ decline in eGFR (compared with sulfonylureas). We did not find evidence for differences in the other study endpoints over the two year study period.

AUTHOR AFFILIATIONS

¹Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK

²Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

³Diabetes Trials Unit, The Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Headington, Oxford, UK

⁴The Comparative Health Outcomes, Policy and Economics (CHOICE) Institute, University of Washington School of Pharmacy, Seattle, WA, USA

⁵Centre for Longitudinal Studies, UCL Social Research Institute, University College London, London, UK

⁶Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

⁷Patient author, Patient Research Champion Team, National Institute for Health and Care Research, London, UK

⁸Diabetes Research Centre, University of Leicester, Leicester, UK

We acknowledge and appreciate the support of the PERMIT study advisory team (Ken Paterson, Jim Lewsey, Ewan Pearson, Sarah Finer, Stephen Evans, Bill Huston, and Rahul Mohan) and the management and administrative support provided by Elizabeth Silver, Anna Carnegie, and Paula Fry.

This study is based in part on data from the Clinical Practice Research Datalink obtained under licence from the UK Medicines and Healthcare Products Regulatory Agency (MHRA). The data are provided by patients and collected by the NHS as part of patient care and support. The interpretation and conclusions contained in this study are those of the authors alone.

Contributors: PB, DGLP, OC, SON, AB, KK, AIA, and RG conceived and designed the study. All authors critically appraised and contributed to the study design and analysis plan. PB requested and extracted the data. PB, DGLP, OC, and SON performed the data management and analysis. PB, DGLP, OC, SON, and RG wrote the manuscript with contributions from all authors. All authors critically revised and approved the manuscript. PB and RG are the guarantors. PB and DGLP share first authorship and contributed equally. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: This work was funded by the National Institute for Health and Care Research (NIHR, grant No NIHR128490). KK is supported by the NIHR Applied Research Collaboration East Midlands (ARC EM)

and the NIHR Leicester Biomedical Research Centre. The funder had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: support from the National Institute for Health and Care Research (NIHR). AIA receives salary from the NIHR Biomedical Research Centre via the Oxford Centre for Diabetes, Endocrinology and Metabolism. AB is an economic advisor on the DIRECT trial, with ongoing responsibility for economic analysis during the long term follow-up phase. He serves as a consultant to Salutis Consulting on projects not related to diabetes. JWB has acted as a consultant to AstraZeneca, Bayer, Novartis, and Roche. DN is the UK Kidney Association Director of Informatics Research; she previously was involved in two GSK funded studies in sub-Saharan Africa unrelated to this work. PC sat on an NIHR Health Technology Assessment commissioning committee until September 2021. AHB has acted as consultant to several pharmaceutical companies within the past three years: AstraZeneca, Boehringer Ingelheim, Daiichi Sankyo, Eli Lilly, Gilead, Idorsia, Novo Nordisk, Rhythm, Roche, and Sanofi. IID has unrestricted research grants from and shares in GSK and a research grant from AstraZeneca. KK has acted as a consultant and speaker or received grants for investigator initiated studies for AstraZeneca, Bayer, Novartis, Novo Nordisk, Sanofi-Aventis, Lilly and Merck Sharp and Dohme, Boehringer Ingelheim, Oramed Pharmaceuticals, Roche, and Applied Therapeutics.

Ethical approval: This research was approved by the London School of Hygiene and Tropical Medicine ethics committee (reference 21 395) and the Independent Scientific Advisory Committee for Medicines and Healthcare Products Regulatory Agency database research (reference 20_064). Although individual consent was not obtained, general practices opt-in to sharing their data for research and individual people can opt-out of sharing their data for research.

Data sharing: Owing to data sharing restrictions, the data used in this study cannot be shared directly. Researchers may, however, apply to use Clinical Practice Research Datalink (CPRD) data linked with other health datasets. Further instructions are available on the CPRD website (<https://cprd.com/>). Codelists to create exposure, outcome, and covariates are published on LSHTM DataCompass (<https://datacompass.lshtm.ac.uk/id/eprint/3743/>).

Transparency: The lead authors (PB and DGLP) and the manuscript's guarantors (PB and RG) affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned have been explained.

Dissemination to participants and related patient and public communities: After publication, the results will be disseminated to policy makers, including those at the National Institute for Health and Care Excellence. We have shared preliminary methodology and results related to this work with two patient panels. With the help of our patient and public involvement and clinical colleagues, we will disseminate the study results after publication with patients, the public, and healthcare professionals through press releases, media communications, social media postings, and presentations at scientific conferences.

Provenance and peer review: Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

- 1 Saeedi P, Petersohn I, Salpea P, et al, IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract* 2019;157:107843. doi:10.1016/j.diabres.2019.107843
- 2 Rodriguez-Gutierrez R, McCoy RG. Measuring What Matters in Diabetes. *JAMA* 2019;321:1865-6. doi:10.1001/jama.2019.4310
- 3 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). *Lancet* 1998;352:837-53. doi:10.1016/S0140-6736(98)07019-6
- 4 Rodriguez-Gutierrez R, Gonzalez-Gonzalez JG, Zuñiga-Hernandez JA, McCoy RG. Benefits and harms of intensive glycemic control in patients with type 2 diabetes. *BMJ* 2019;367:l5887. doi:10.1136/bmj.l5887

- 5 American Diabetes Association. 6. Glycemic Targets: *Standards of Medical Care in Diabetes-2020*. *Diabetes Care* 2020;43(Suppl 1):S66-76. doi:10.2337/dc20-S006
- 6 National Institute for Health and Care Excellence. Canagliflozin, dapagliflozin and empagliflozin as monotherapies for treating type 2 diabetes. Technology appraisal guidance TA390. NICE, 2016. <https://www.nice.org.uk/guidance/ta390>.
- 7 National Institute for Health and Care Excellence. Type 2 diabetes in adults: management. NICE guideline NG28. NICE, 2022. <https://www.nice.org.uk/guidance/ng28/chapter/Recommendations#reviewing-drug-treatments>.
- 8 Scottish Intercollegiate Guidelines Network. SIGN 154: Pharmacological management of glycaemic control in people with type 2 diabetes. SIGN, 2017. <https://www.sign.ac.uk/our-guidelines/pharmacological-management-of-glycaemic-control-in-people-with-type-2-diabetes/>.
- 9 Davies MJ, Aroda VR, Collins BS, et al. Management of Hyperglycemia in Type 2 Diabetes, 2022. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* 2022;45:2753-86. doi:10.2337/dci22-0034
- 10 Khunti K, Charbonnel B, Cooper A, et al, DISCOVER investigators. Associations between second-line glucose-lowering combination therapies with metformin and HbA1c, body weight, quality of life, hypoglycaemic events and glucose-lowering treatment intensification: The DISCOVER study. *Diabetes Obes Metab* 2021;23:1823-33. doi:10.1111/dom.14400
- 11 Bidulka P, Mathur R, Lugo-Palacios DG, et al. Ethnic and socioeconomic disparities in initiation of second-line antidiabetic treatment for people with type 2 diabetes in England: A cross-sectional study. *Diabetes Obes Metab* 2023;25:282-92. doi:10.1111/dom.14874
- 12 Wilkinson S, Douglas I, Stirnadel-Farrant H, et al. Changing use of antidiabetic drugs in the UK: trends in prescribing 2000-2017. *BMJ Open* 2018;8:e022768. doi:10.1136/bmjopen-2018-022768
- 13 Farmer RE, Beard I, Raza SI, et al. Prescribing in Type 2 Diabetes Patients With and Without Cardiovascular Disease History: A Descriptive Analysis in the UK CPRD. *Clin Ther* 2021;43:320-35. doi:10.1016/j.clinthera.2020.12.015
- 14 Donnan JR, Grandy CA, Chibrikov E, et al. Comparative safety of the sodium glucose co-transporter 2 (SGLT2) inhibitors: a systematic review and meta-analysis. *BMJ Open* 2019;9:e022577. doi:10.1136/bmjopen-2018-022577
- 15 Baigent C, Emberson J, Haynes R, et al, Nuffield Department of Population Health Renal Studies Group/SGLT2 inhibitor Meta-Analysis Cardio-Renal Trialists' Consortium. Impact of diabetes on the effects of sodium glucose co-transporter-2 inhibitors on kidney outcomes: collaborative meta-analysis of large placebo-controlled trials. *Lancet* 2022;400:1788-801. doi:10.1016/S0140-6736(22)02074-8
- 16 Zinman B, Wanner C, Lachin JM, et al, EMPA-REG OUTCOME Investigators. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *N Engl J Med* 2015;373:2117-28. doi:10.1056/NEJMoa1504720
- 17 Neal B, Perkovic V, Mahaffey KW, et al, CANVAS Program Collaborative Group. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. *N Engl J Med* 2017;377:644-57. doi:10.1056/NEJMoa1611925
- 18 Wiviott SD, Raz I, Bonaca MP, et al, DECLARE-TIMI 58 Investigators. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med* 2019;380:347-57. doi:10.1056/NEJMoa1812389
- 19 Cannon CP, Pratley R, Dagogo-Jack S, et al, VERTIS CV Investigators. Cardiovascular Outcomes with Ertugliflozin in Type 2 Diabetes. *N Engl J Med* 2020;383:1425-35. doi:10.1056/NEJMoa2004967
- 20 Perkovic V, Jardine MJ, Neal B, et al, CREDESCENCE Trial Investigators. Canagliflozin and Renal Outcomes in Type 2 Diabetes and Nephropathy. *N Engl J Med* 2019;380:2295-306. doi:10.1056/NEJMoa1811744
- 21 Herrington WG, Staplin N, Wanner C, et al, The EMPA-KIDNEY Collaborative Group. Empagliflozin in Patients with Chronic Kidney Disease. *N Engl J Med* 2023;388:117-27. doi:10.1056/NEJMoa2204233
- 22 McMurray JJV, Solomon SD, Inzucchi SE, et al, DAPA-HF Trial Committees and Investigators. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. *N Engl J Med* 2019;381:1995-2008. doi:10.1056/NEJMoa1911303
- 23 Heerspink HJL, Stefánsson BV, Correa-Rotter R, et al, DAPA-CKD Trial Committees and Investigators. Dapagliflozin in Patients with Chronic Kidney Disease. *N Engl J Med* 2020;383:1436-46. doi:10.1056/NEJMoa2024816
- 24 Packer M, Anker SD, Butler J, et al, EMPEROR-Reduced Trial Investigators. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. *N Engl J Med* 2020;383:1413-24. doi:10.1056/NEJMoa2022190
- 25 Rosenstock J, Perkovic V, Johansen OE, et al, CARMELINA Investigators. Effect of Linagliptin vs Placebo on Major Cardiovascular Events in Adults With Type 2 Diabetes and High Cardiovascular and Renal Risk: The CARMELINA Randomized Clinical Trial. *JAMA* 2019;321:69-79. doi:10.1001/jama.2018.18269
- 26 Green JB, Bethel MA, Armstrong PW, et al, TECOS Study Group. Effect of Sitagliptin on Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med* 2015;373:232-42. doi:10.1056/NEJMoa1501352
- 27 Ferrannini E, Fonseca V, Zinman B, et al. Fifty-two-week efficacy and safety of vildagliptin vs. glimepiride in patients with type 2 diabetes mellitus inadequately controlled on metformin monotherapy. *Diabetes Obes Metab* 2009;11:157-66. doi:10.1111/j.1463-1326.2008.00994.x
- 28 Nauck MA, Meininger G, Sheng D, Terranella L, Stein PP, Sitagliptin Study 024 Group. Efficacy and safety of the dipeptidyl peptidase-4 inhibitor, sitagliptin, compared with the sulfonylurea, glipizide, in patients with type 2 diabetes inadequately controlled on metformin alone: a randomized, double-blind, non-inferiority trial. *Diabetes Obes Metab* 2007;9:194-205. doi:10.1111/j.1463-1326.2006.00704.x
- 29 Nathan DM, Lachin JM, Balasubramanyam A, et al, GRADE Study Research Group. Glycemia Reduction in Type 2 Diabetes - Glycemic Outcomes. *N Engl J Med* 2022;387:1063-74. doi:10.1056/NEJMoa2200433
- 30 Rosenstock J, Kahn SE, Johansen OE, et al, CAROLINA Investigators. Effect of Linagliptin vs Glimepiride on Major Adverse Cardiovascular Outcomes in Patients With Type 2 Diabetes: The CAROLINA Randomized Clinical Trial. *JAMA* 2019;322:1155-66. doi:10.1001/jama.2019.13772
- 31 Leiter LA, Yoon KH, Arias P, et al. Canagliflozin provides durable glycemic improvements and body weight reduction over 104 weeks versus glimepiride in patients with type 2 diabetes on metformin: a randomized, double-blind, phase 3 study. *Diabetes Care* 2015;38:355-64. doi:10.2337/dc13-2762
- 32 D'Andrea E, Wexler DJ, Kim SC, Paik JM, Alt E, Paterno E. Comparing Effectiveness and Safety of SGLT2 Inhibitors vs DPP-4 Inhibitors in Patients With Type 2 Diabetes and Varying Baseline HbA1c Levels. *JAMA Intern Med* 2023;183:242-54. doi:10.1001/jamainternmed.2022.6664
- 33 Wilkinson S, Williamson E, Pokrajac A, et al. Comparative effects of sulphonylureas, dipeptidyl peptidase-4 inhibitors and sodium-glucose co-transporter-2 inhibitors added to metformin monotherapy: a propensity-score matched cohort study in UK primary care. *Diabetes Obes Metab* 2020;22:847-56. doi:10.1111/dom.13970
- 34 Khunti K, Kosiborod M, Kim DJ, et al, CVD-REAL Investigators and Study Group. Cardiovascular outcomes with sodium-glucose cotransporter-2 inhibitors vs other glucose-lowering drugs in 13 countries across three continents: analysis of CVD-REAL data. *Cardiovasc Diabetol* 2021;20:159. doi:10.1186/s12933-021-01345-z
- 35 Deng Y, Polley EC, Wallach JD, et al. Emulating the GRADE trial using real world data: retrospective comparative effectiveness study. *BMJ* 2022;379:e070717. doi:10.1136/bmj-2022-070717
- 36 Nathan DM, Lachin JM, Bebu I, et al, GRADE Study Research Group. Glycemia Reduction in Type 2 Diabetes - Microvascular and Cardiovascular Outcomes. *N Engl J Med* 2022;387:1075-88. doi:10.1056/NEJMoa2200436
- 37 Wexler DJ, de Boer IH, Ghosh A, et al, GRADE Research Group. Comparative Effects of Glucose-Lowering Medications on Kidney Outcomes in Type 2 Diabetes: The GRADE Randomized Clinical Trial. *JAMA Intern Med* 2023;183:705-14. doi:10.1001/jamainternmed.2023.1487
- 38 Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med* 2014;33:2297-340. doi:10.1002/sim.6128
- 39 Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *Int J Biostat* 2007;3:14. doi:10.2202/1557-4679.1072
- 40 Xie Y, Bowe B, Xian H, Loux T, McGill JB, Al-Aly Z. Comparative effectiveness of SGLT2 inhibitors, GLP-1 receptor agonists, DPP-4 inhibitors, and sulfonylureas on risk of major adverse cardiovascular events: emulation of a randomised target trial using electronic health records. *Lancet Diabetes Endocrinol* 2023;11:644-56. doi:10.1016/S2213-8587(23)00171-7
- 41 Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol* 2016;183:758-64. doi:10.1093/aje/kwv254
- 42 Bidulka P, O'Neill S, Basu A, et al. Protocol for an observational cohort study investigating personalised medicine for intensification of treatment in people with type 2 diabetes mellitus: the PERMIT study. *BMJ Open* 2021;11:e046912. doi:10.1136/bmjopen-2020-046912

- 43 Bidulka P, Lugo-Palacios DG, Carroll O, O'Neill S, Grieve R. Statistical Analysis Plan (SAP): Personalised Medicine for Intensification of Treatment (PERMIT) study, Version 1.0, July 2023. <https://www.lshtm.ac.uk/media/72276>.
- 44 Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;44:827-36. doi:10.1093/ije/dyv098
- 45 Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol* 2019;48:1740-1740g. doi:10.1093/ije/dyz034
- 46 Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017;46:1093-1093i. doi:10.1093/ije/dyx015
- 47 Herrett E, Shah AD, Boggon R, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ* 2013;346:f2350. doi:10.1136/bmj.f2350
- 48 Bidulka P, Scott J, Taylor DM, et al. Impact of chronic kidney disease on case ascertainment for hospitalised acute myocardial infarction: an English cohort study. *BMJ Open* 2022;12:e057909. doi:10.1136/bmjopen-2021-057909
- 49 Linked HES-ONS mortality data. NHS Digital; 2020. <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/linked-hes-ONS-mortality-data#ons-mortality-data>.
- 50 Penney B. The English Indices of Deprivation 2019. 2019.
- 51 Wilkinson S, Douglas IJ, Williamson E, et al. Factors associated with choice of intensification treatment for type 2 diabetes after metformin monotherapy: a cohort study in UK primary care. *Clin Epidemiol* 2018;10:1639-48. doi:10.2147/CLEPS176142
- 52 NHS England - About Us - Regional Teams. NHS England. <https://www.england.nhs.uk/about/regional-area-teams/>.
- 53 Bhaskaran K, Forbes HJ, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). *BMJ Open* 2013;3:e003389. doi:10.1136/bmjopen-2013-003389
- 54 Levey AS, Inker LA, Matsushita K, et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *Am J Kidney Dis* 2014;64:821-35. doi:10.1053/j.ajkd.2014.07.030
- 55 Levin A, Agarwal R, Herrington WG, et al, participant authors of the International Society of Nephrology's 1st International Consensus Meeting on Defining Kidney Failure in Clinical Trials. International consensus definitions of clinical trial outcomes for kidney failure: 2020. *Kidney Int* 2020;98:849-59. doi:10.1016/j.kint.2020.07.013
- 56 Staiger D, Stock JH. Instrumental Variables Regression with Weak Instruments. *Econometrica* 1997;65:557-86doi:10.2307/2171753.
- 57 Moler-Zapata S, Grieve R, Basu A, O'Neill S. How does a local instrumental variable method perform across settings with instruments of differing strengths? A simulation study and an evaluation of emergency surgery. *Health Econ* 2023;32:2113-26. doi:10.1002/hec.4719
- 58 Swanson SA, Miller M, Robins JM, Hernán MA. Definition and evaluation of the monotonicity condition for preference-based instruments. *Epidemiology* 2015;26:414-20. doi:10.1097/EDE.0000000000000279
- 59 Vytlacil E. Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica* 2002;70:331-41doi:10.1111/1468-0262.00277.
- 60 Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *J Health Econ* 2008;27:531-43. doi:10.1016/j.jhealeco.2007.09.009
- 61 Basu A, Coe NB, Chapman CG. 2SLS versus 2SRI: Appropriate methods for rare outcomes and/or rare exposures. *Health Econ* 2018;27:937-55. doi:10.1002/hec.3647
- 62 Frank IE, Friedman JH. A Statistical View of Some Chemometrics Regression Tools. *Technometrics* 1993;35:109-35doi:10.1080/00401706.1993.10485033.
- 63 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc B* 1996;58:267-88doi:10.1111/j.2517-6161.1996.tb02080.x.
- 64 Belloni A, Chen D, Chernozhukov V, Hansen C. Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. *Econometrica* 2012;80:2369-429doi:10.3982/ECTA9626.
- 65 Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, 1987doi:10.1002/9780470316696.
- 66 van Buuren S, Oudshoorn CGM. Multivariate Imputation by Chained Equations: MICE V1.0 User's manual. TNO Report PG/VGZ/00.038. Leiden, Netherlands; 2000.
- 67 Powney M, Williamson P, Kirkham J, Kolamunnage-Dona R. A review of the handling of missing longitudinal outcome data in clinical trials. *Trials* 2014;15:237. doi:10.1186/1745-6215-15-237
- 68 Lee KJ, Roberts G, Doyle LW, Anderson PJ, Carlin JB. Multiple imputation for missing data in a longitudinal cohort study: a tutorial based on a detailed case study involving imputation of missing outcome data. *Int J Soc Res Methodol* 2016;19:575-91doi:10.1080/13645579.2015.1126486.
- 69 Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Med Res Methodol* 2014;14:75. doi:10.1186/1471-2288-14-75
- 70 White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med* 2009;28:1982-98. doi:10.1002/sim.3618
- 71 Bartlett JW, Hughes RA. Bootstrap inference for multiple imputation under unconfounding and misspecification. *Stat Methods Med Res* 2020;29:3533-46. doi:10.1177/0962280220932189
- 72 Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. *Stat Med* 2018;37:2252-66. doi:10.1002/sim.7654
- 73 Therneau T. A Package for Survival Analysis in R: R package version 3.4-0, 2022. <https://CRAN.R-project.org/package=survival>.
- 74 van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw* 2011;45:1-67doi:10.18637/jss.v045.i03.
- 75 Team RCR. *A language and environment for statistical computing*. R Foundation for Statistical Computing, 2022, <https://www.R-project.org/>.
- 76 StataCorp. *Stata Statistical Software: Release 17 College Station*. StataCorp LLC, 2021.
- 77 Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. *BMJ* 2019;367:l5657. doi:10.1136/bmj.l5657
- 78 Wooldridge JM. Inverse probability weighted estimation for general missing data problems. *J Econom* 2007;141:1281-301doi:10.1016/j.jeconom.2007.02.002.
- 79 Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value Health* 2010;13:273-7. doi:10.1111/j.1524-4733.2009.00671.x
- 80 Stürmer T, Webster-Clark M, Lund JL, et al. Propensity Score Weighting and Trimming Strategies for Reducing Variance and Bias of Treatment Effect Estimates: A Simulation Study. *Am J Epidemiol* 2021;190:1659-70. doi:10.1093/aje/kwab041
- 81 Stürmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution—a simulation study. *Am J Epidemiol* 2010;172:843-54. doi:10.1093/aje/kwq198
- 82 Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. *Am J Epidemiol* 2011;173:761-7. doi:10.1093/aje/kwq439
- 83 PERMIT. Investigating the effectiveness of different treatments for Type 2 Diabetes Mellitus, to help choose the right treatments for patients. London School of Hygiene and Tropical Medicine, 2023. <https://www.lshtm.ac.uk/research/centres-projects-groups/permit#welcome>.
- 84 Hansford HJ, Cashin AG, Jones MD, et al. Reporting of Observational Studies Explicitly Aiming to Emulate Randomized Trials: A Systematic Review. *JAMA Netw Open* 2023;6:e2336023. doi:10.1001/jamanetworkopen.2023.36023
- 85 Wang SV, Schneeweiss S, Franklin JM, et al, RCT-DUPLICATE Initiative. Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses: Results of 32 Clinical Trials. *JAMA* 2023;329:1376-85. doi:10.1001/jama.2023.4221
- 86 National Institute for Health and Care Excellence. Type 2 diabetes [B] Pharmacological therapies with cardiovascular and other benefits in people with type 2 diabetes, NICE guideline NG28. NICE, 2022. <https://www.nice.org.uk/guidance/ng28/evidence/b-pharmacological-therapies-with-cardiovascular-and-other-benefits-in-people-with-type-2-diabetes-pdf-10956473392>.
- 87 Sim R, Chong CW, Loganadan NK, et al. Comparative effectiveness of cardiovascular, renal and safety outcomes of second-line antidiabetic drugs use in people with type 2 diabetes: A systematic review and network meta-analysis of randomised controlled trials. *Diabet Med* 2022;39:e14780. doi:10.1111/dme.14780
- 88 Cavender MA, Norhammar A, Birkeland KI, et al, CVD-REAL Investigators and Study Group. SGLT-2 Inhibitors and Cardiovascular Risk: An Analysis of CVD-REAL. *J Am Coll Cardiol* 2018;71:2497-506. doi:10.1016/j.jacc.2018.01.085
- 89 Ni L, Yuan C, Chen G, Zhang C, Wu X. SGLT2i: beyond the glucose-lowering effect. *Cardiovasc Diabetol* 2020;19:98. doi:10.1186/s12933-020-01071-y

Supplementary information: Supplementary methods, tables, and figures