Statistical Analysis Plan (SAP)

PERsonalised Medicine for Intensification of Treatment (PERMIT) study

Version 1.0, July 2023

Prepared by:

Patrick Bidulka, David G Lugo-Palacios, Orlagh Carroll, Stephen O'Neill, and Richard Grieve

With input from the broader PERMIT team:

Amanda I Adler, Anirban Basu, Richard J Silverwood, Jonathan W. Bartlett, Dorothea Nitsch, Paul Charlton, Andrew H Briggs, Liam Smeeth, Ian J Douglas, and Kamlesh Khunti

Contents

1. Overview of Statistical Analysis Plan (SAP) 3

2. Background to the PERMIT study 4

3. Target Trial Emulation 6

3.1 Data sources 6

3.2 Study population 7

3.3 Sample size 9

3.4 Treatments of interest 10

3.5 Outcomes 11

3.6 Potential confounders/covariables 16

4. Analysis 19

4.1 The key IV assumptions 19

4.2 Checking IV assumptions 21

4.3 Accounting for defiers/always takers 21

4.4 Two stage Least Squares (2SLS) 21

4.5 Two stage Residual Inclusion (2SRI) 22

4.6 Subgroup analyses 24

4.7 Handling missing data, censoring and loss to follow-up 25

4.8 Alternative analyses 31

Appendix 33

6. References 39

1. Overview of Statistical Analysis Plan (SAP)

This document sets out the proposed presentation and analysis for the main paper(s) reporting the results from main elements of the PERMIT study that will assess the relative effectiveness of second-line antidiabetic drug treatments added on to metformin in type 2 diabetes from individual patient data extracted from the Clinical Practice Research Datalink (CPRD) and linked datasets. This document should be read in conjunction with the study protocol,[1] and the preceding description of the study cohort.[2] The purpose of the SAP is to pre-specify the main analyses that will be reported in subsequent papers. Any subsequent exploratory analyses will not be bound by this strategy. Suggestions for subsequent analyses by oversight committees, journal editors or referees, will be considered carefully in line with the principles of this analysis plan. Any deviations from the statistical analysis plan will be described and justified in the final report to the funder.

The SAP is structured as follows: in Section 2, we provide an overview of the PERMIT study, and describe the main objectives pertaining to this SAP. In Section 3, we describe the data sources and the main 'standpoints' of the PERMIT 'target trial emulation'. A target trial is a hypothetical pragmatic RCT for assessing comparative effectiveness using observational data that is designed to minimise prognostic differences between the comparison groups.[3,4] The target trial framework requires the study to define the main elements of the target trial's protocol, including eligibility criteria, and the treatment strategies. The PERMIT target trial will apply eligibility criteria to identify people with Type 2 diabetes mellitus (T2DM) of similar prognosis prior to initiating second-line antidiabetic drug treatments added on to metformin but different treatment assignments, and will compare their outcomes over the subsequent 2 years. In Section 4, we describe the proposed analyses, and in particular detail

the instrumental variable analysis, which aims to provide accurate estimates of treatment effectiveness even when there are residual baseline prognostic differences between the comparison groups.[5] In section 5, we outline the main tables and figures that we anticipate presenting in the main paper and supplementary appendices.

2. Background to the PERMIT study

People living with T2DM are usually prescribed metformin monotherapy as first-line oral antidiabetic treatment to control blood glucose, after blood glucose fails to be controlled by lifestyle interventions.[6] The National Institute of Health and Care Excellence (NICE) then recommends several glucose lowering drug classes, added on to metformin, as second-line oral antidiabetic treatment if metformin monotherapy inadequately controls blood glucose and the person tolerates metformin monotherapy. These drug classes include sulfonylureas (SU), dipeptidyl peptidase-4 inhibitors (DPP4i), and, for patients who are at significant risk of hypoglycaemia or for whom sulfonylurea is contraindicated, sodium-glucose co-transporter 2 inhibitors (SGLT2i).[6] In 2022, NICE updated guidelines to recommend SGLT2i in combination with metformin for those with a high risk of, or with, cardiovascular disease; however, for all other people, any of these 3 treatments may be suitable.[6] Similarly, international guidance and consensus reports recommend SGLT2i for people with established atherosclerotic CVD, heart failure, and chronic kidney disease (CKD).[7] These guidelines/guidance draw on evidence from placebo-controlled randomised controlled trials (RCTs) showing improved CVD and kidney disease outcomes in people prescribed SGLT2i.

The GRADE study, published in 2022,[8-10] compared glycaemic control and cardiovascular and kidney endpoints between SU, DPP4i, glucagon-like peptide-1 receptor agonist (GLP1-RA),

and insulin. However, there are no published head-to-head RCTs that directly compare SU, DPP4i, and SGLT2i. A network meta-analysis published in 2022 by NICE[11] compared cardiovascular outcomes between several classes of antidiabetic medicines using fifteen placebo-controlled RCTs, and one trial with a direct comparison (DPP4i vs SU).[12] The study reported that empagliflozin (SLGT2i drug class) and subcutaneous semaglutide (glucagon-like peptide-1 receptor agonist (GLP1-RA)) reduced all-cause and cardiovascular mortality compared to other drug classes, and to placebo. The network meta-analysis also found that several SGLT2i drugs reduced rates of heart failure hospitalisation, while only semaglutide reported clinically meaningful benefits in terms of reduced rates of major adverse cardiovascular event (MACE), according to the composite three-point scale.

While this network meta-analysis[11] and placebo-controlled trials indicate that of the 3 classes of second-line antidiabetic drug treatments included in this study that are most commonly prescribed in the UK, SGLT2i may be the most effective in preventing cardiovascular outcomes, there is a lack of head-to-head trial evidence that directly assesses which drug class is most effective at controlling blood glucose levels, and reducing the incidence of the complications of diabetes among people with T2DM. Current NICE guidelines do not recommend a specific class of second-line antidiabetic drug treatment for people not at high-risk or with cardiovascular disease. Previous research has highlighted the wide variation in second-line oral antidiabetic treatment prescribing in the United Kingdom.[13]

The PERMIT study aims to use this variation to conduct a natural experiment – that is, using an instrumental variable approach[5] to find a 'natural randomiser' – using routinely collected health data in England to compare the relative effectiveness and cost-effectiveness of SU,

DPP4i, and SGLT2i as second-line oral antidiabetic treatment, overall, and among clinically-defined subgroups. A major challenge for estimating treatment effectiveness from routine data is the potential for confounding by indication that may arise due to measured or unmeasured prognostic differences (e.g., different levels of glycaemic control) that may arise between the treatment groups being compared, prior to starting second-line treatment. We will address this challenge in two ways. First, we will follow the principles of target trial emulation, by applying consistent eligibility criteria across the comparators of interest defining treatment strategies and outcomes, follow-up and the causal contrasts of interest.[4,14] Second, we will undertake an instrumental variable (IV) analysis, as this can provide accurate estimates of treatment effectiveness even when there are unmeasured differences between the comparison groups.[5,15] We will follow previous studies in using variation in provider prescribing as an instrumental variable to compare outcomes across patients receiving second-line treatments added on to metformin with these 3 classes.[15] The next sections detail the data sources that the study will use, before outlining the target trial principles that we will follow.

3. Target Trial Emulation

3.1 Data sources

Primary care data

We will use primary care data from the Clinical Practice Research Datalink (CPRD) Gold and Aurum datasets, which includes approximately 20% of the UK population.[16,17] General Practices (GP) in the United Kingdom (UK) using Vision and EMIS software contribute to the

CPRD Gold and Aurum datasets, respectively. CPRD data include longitudinal information on primary care diagnoses, prescriptions, demographic information, and laboratory test results.

Secondary care data

We will link primary care CPRD data to Hospital Episode Statistics (HES) admitted patient care (APC) data, which includes information related to all in-patient hospitalisations in England funded by the National Health Service (NHS), and reports admission and discharge dates, diagnoses, and demographic information.[18] Our primary care population will therefore be restricted to those registered at GP practices in England.

Other health data

We will link the CPRD-HES data to mortality data from England for the Office of National Statistics (ONS) and the patient-level Index of Multiple Deprivation (IMD).[19,20] ONS data include the date and cause of death date (International Classification of Diseases 10th edition (ICD-10) codes) for all deaths registered in England and Wales. The IMD is a composite score which ranks individuals' deprivation according to the local area of their place of residence. IMD scores will be reported in quintiles (from 1, resident in the least deprived, to 5 the most deprived local area).

3.2 Study population

In line with target trial principles, we will define the study population according to eligibility criteria, which have to be met prior to 'time zero' (i.e. 'baseline') and is the analogue to the time of randomisation in an RCT. In the PERMIT study, time zero is defined as the time of the first prescription for any of the second-line treatments (see section 2.4).

We will include people with T2DM who intensified from first- to second-line oral antidiabetic treatment between 1 January 2015 and 31 December 2020. We choose to start the study in 2015, since SGLT2i were rarely used prior to 2014,[13] and we require 1 year of prescribing history to define the instrument in our instrumental variable analysis. The PERMIT study population will be defined by people with T2DM who were registered with GPs in England, as the instrumental variable is the previous prescribing history of groups of GPs defined at the level of the Clinical Commissioning Group (CCG) (now Integrated Care Systems). The inclusion criteria also include those individuals whose primary care data were linked to HES/ONS/IMD data to enable the requisite outcomes to be reported (see section 2.5. Outcomes). Detailed inclusion/exclusion criteria are presented in **Table 1.**

Table 1: Inclusion and exclusion criteria for the PERMIT study population

Inclusion criteria	**Exclusion criteria**
• Aged ≥18 years (the study is of adults only). • T2DM diagnosis code, to avoid including people prescribed antidiabetic drugs for other indications (e.g., polycystic ovarian syndrome). • Prescribed metformin monotherapy as first-line oral antidiabetic treatment, on the same day or following a T2DM diagnosis. • Registered with GP in England with acceptable data standards flag by CPRD (to help ensure adequate data availability). • Registered with GP for ≥1 year prior to first metformin prescription (to help ensure adequate baseline data availability and reduce recording of past events as incident).	• Prescribed >1 non-metformin antidiabetic drugs on the date of second-line treatment initiation (beyond study scope). • Initiates second-line oral antidiabetic treatment with drug class other than SU, DPP4i, or SGLT2i (beyond study scope). • Latest eGFR recorded by the GP is <30mL/min/1.73m^2 (since at the time of data-collection most GPs would not have prescribed metformin for people with eGFR <30ml/min; the results from the DAPA-CKD trial (which did randomise people with eGFR less than 30ml/min/1.73m^2) were only available towards the very end of the study period and are unlikely to have informed decisions taken in primary care).

• Initiate SU, DPP4i, or SGLT2i between 1 January 2015 to 31 December 2020 (the study period). • At least 1 metformin prescription within 60 days prior to new second-line drug, and at least one metformin prescription on the same day or within 60 days after new second-line drug, to ensure the person is adding on to metformin and not switching. • Linked to HES/ONS/IMD data (to help ensure outcomes are captured).	• Women who have a record of pregnancy in primary care within 1 year prior to second-line antidiabetic treatment initiation (since guidelines are different for this group).

3.3 Sample size

The effect of interest is the absolute difference between the change in haemoglobin A1c (HbA1c) from baseline to 12 months under a treatment (A) versus the change under the comparator treatment (B). Power calculations were conducted prior to accessing study data, recognising the Food & Drug Administration (FDA) recommendation that an average between-treatment absolute difference in HbA1c from baseline to 12 months of 4.4 mmol/mol (0.4 percentage points) is of clinical significance,[21] and assuming a standard deviation for HbA1c at baseline of 14 mmol/mol (1.3 percentage points).[22] This difference will be calculated as: (HbA1c at 12 months –baseline for treatment B) – (HbA1c at 12 months – baseline for treatment A). We follow methodological recommendations for power calculations with IV designs and consider that the IV predicts 80% of those prescribed each second-line treatment, but also consider scenarios where the IV is weaker (70% compliance) and stronger (90% compliance).[23] We require 80% power at the 5% (2-

sided) level of statistical significance, but conservatively base the power calculations on a type I error corrected for the 3 comparison (I.e. use 5% / 3 as the significance level).

Table 2 shows the requisite sample sizes for the 2 treatment groups projected to have the fewest participants (SU and SGLT2i). Note that while there is only a single primary outcome (HbA1C) and timepoint (12 months), there are several secondary endpoints and timepoints of interest. Rather than formally adjust for these multiple comparisons, we will recognise the 'hypothesis generating' nature of these comparisons in the interpretation of the results.[23]

Table 2: Required sample size (N) for the IV design according to instrument strength (level of compliance) and magnitude of effect size at 80% power and 5% (2-sided) level of statistical significance

	Level of compliance (IV strength)					
	70%		**80%**		**90%**	
Effect size: between-treatment difference in mean HbA1c reduction baseline to 12 months in mmol/mol and percentage points(%)	**SU**	**SGLT2i**	**SU**	**SGLT2i**	**SU**	**SGLT2i**
3 mmol/mol (0.3 %)	4556	1952	3488	1495	2756	1181
4 mmol/mol (0.4 %)	2563	1098	1962	841	1550	664
5 mmol/mol (0.5 %)	1640	703	1256	538	992	425

3.4 Treatments of interest

We will compare (1) DPP4i versus SU, (2) SGLT2i versus SU, and (3) SGLT2i vs DPP4i all as second-line oral antidiabetic treatment and as add-ons to metformin. Although the latter are starting to be used more following the 2022 NICE guidelines update,[6] our study period ends in 2021.

This study will use an intention-to-treat (ITT) approach, whereby each person will contribute follow-up to the original exposure group to which they were assigned, irrespective of which treatments they may be prescribed subsequently. To help with the interpretation of the ITT approach, we will report second-line treatment duration during follow-up. We will also report the proportion of each generic drug prescribed within each of the 3 drug classes.

Second-line oral antidiabetic treatment initiation (first-stage intensification) will be the date of the first prescription for SU, DPP4i, or SGLT2i following metformin monotherapy (hereafter referred to as the 'baseline' date). To ensure that people are adding on to metformin monotherapy, and not switching, we will require at least one metformin prescription on the same day or within 60 days after the first prescription for the second-line drug.[24,25] We will consider the potential impact of immortal time bias, which arises because the individual must remain in the sample for at least 60 days for it to be determined whether they meet this criteria, in a sensitivity analysis (see section 3.7. Sensitivity analyses). The bias occurs as some events (e.g. censoring by death) cannot have occurred in this time window.

3.5 Outcomes

The primary outcome will be absolute change in HbA1c, reported in mmol/mol, at 12 months follow-up recorded in CPRD between groups. This change in HbA1c will be quantified by contrasting follow-up versus baseline laboratory test data recorded on CPRD for each comparison group. We will use a window of ±90 days (3 months) around the 12-month follow-up date for the primary outcome since pilot work showed substantial missingness in the recording of HbA1c at this time point. If there are more than one HbA1c measures recorded during the outcome window, we will use the HbA1c measure closest to

the 12-month follow-up date. Patients without the relevant measurement between 9 and 15 months will be designated as having 'missing 12-month data' (see section 3.6. Missing data). For the baseline measure, we will use the most recent HbA1c prior to the baseline date (second-line treatment initiation), excluding any measures which were recorded over 180 days prior to baseline. Those without an HbA1c measure within 180 days prior to baseline will be considered as having missing baseline HbA1c (see section 3.6. Missing data).

Secondary outcomes that we will also investigate at 12-months post-baseline are: estimated mean difference between the comparison groups in the absolute change from baseline: in glomerular filtration rate (eGFR), body mass index (BMI), and systolic blood pressure (SBP). We will use the same strategy as for the primary outcome to define follow-up outcome measures for these secondary outcomes. However, as previous studies have suggested that in routine primary care data, these outcomes are measured less frequently than HbA1c, we will allow for baseline measures up to 540 days prior to the baseline timepoint (time zero).[25] For all continuous outcomes (HbA1c, eGFR, BMI, SBP), we will also report changes from baseline at multiple time points 6-, 24-, 36-, 48-, and 60-months follow-up, with the main presentation of results at 12- and 24- months. For each of the follow-up timepoints we will only consider observations within three months of the designated timepoint, and if no such observations are available the measure will be designated as 'missing' or 'censored'.

Additional secondary outcomes will be in the form of 'time to first event' and will include: (a) a 40% decline in eGFR from baseline, which could be a marker of the rarer outcome end-stage kidney disease,[26] (b) a major adverse kidney event (MAKE), a composite outcome for the earliest of a decline in eGFR from baseline of 40%, end-stage kidney disease (ESKD), and all-cause mortality,[27] (c) heart failure hospitalisation, (d) 3-point major adverse

cardiovascular event (MACE), a composite outcome for the earliest of myocardial infarction (MI), stroke, and CVD death, (e) MI, (f) Stroke, and (g) all-cause mortality. Details to define each of these outcomes are found in **Table 3.**

Table 3: List of primary and secondary outcomes and the data sources used to define them

Type of outcome	**Outcome**	**Data source**	**Details**
Continuous	Absolute change in HbA1c	CPRD	CPRD Laboratory measures of HbA1c
Continuous	Absolute change in eGFR	CPRD	CPRD Laboratory measures of serum creatinine, converted to eGFR using the 2009 CKD-EPI equation without adjustment for ethnicity
Continuous	Absolute change in BMI	CPRD	CPRD Measures of body weight and height, using a previously developed algorithm to define BMI in CPRD data[28]
Continuous	Absolute change in SBP	CPRD	CPRD Measures of systolic blood pressure
Time-to-event	MACE, including MI, stroke, and all-cause mortality	CPRD, HES, ONS	CPRD Diagnosis codes for MI, stroke HES Diagnosis codes for MI, stroke in the first or second diagnostic position of any episode in a spell) ONS Death date
Time-to-event	MI	CPRD, HES	CPRD Diagnosis code for MI HES Diagnosis code for MI in the first or second diagnostic position of any episode in a spell
Time-to-event	Stroke	CPRD, HES	CPRD Diagnosis code for stroke HES

Type of outcome	Outcome	Data source	Details
			Diagnosis code for stroke in the first or second episode of any episode in a spell
Time-to-event	All-cause mortality	ONS	ONS Death date
Time-to-event	Heart failure hospitalisation	HES	HES Diagnosis code for HF in the first or second diagnostic position of any episode in a spell
Time-to-event	Major adverse kidney event (MAKE), a composite outcome including 40% decline in eGFR, ESKD, and all-cause mortality	CPRD, ONS	CPRD 40% decline in eGFR at baseline (using eGFR derived from laboratory measures of serum creatinine) ESKD (clinical codes diagnosing ESKD/chronic dialysis/kidney transplant) ONS Death date
Time-to-event	40% decline in eGFR from baseline, which could be a proxy for the rarer ESKD outcome[26]	CPRD	CPRD Laboratory measures for serum creatinine to derive eGFR using the 2009 CKD-EPI equation without correction for ethnicity
Time-to-event	ESKD		CPRD Clinical codes for diagnosis of ESKD, or dialysis/kidney transplant codes

Other secondary outcomes will be exploratory, and are not planned for inclusion in the main paper. These include clinical measures from the blood (High Density Lipoprotein (HDL), Low Density Lipoprotein (LDL), total cholesterol, triglycerides, diastolic blood pressure (DBP)) analysed as continuous outcomes, and additional macro- and micro-vascular outcomes (unstable angina, hypoglycaemia, nephropathy, and lower limb amputation).

We will describe process measures including: time to cessation of second-line antidiabetic treatment, the proportion of people who stop their second-line treatment, and those who switch to a further antidiabetic drug.

Time-to-event (TTE) outcomes will be defined in CPRD data using Read codes (CPRD Gold), SNOMED codes (CPRD Aurum), or laboratory measures captured in the primary care record. Outcomes will be defined in HES data using ICD-10 codes, restricting to those diagnoses in the first or second diagnostic position of any episode during a single hospitalisation (referred to as a "spell"), to ensure we are capturing the incident outcome event (usually recorded in the first or second diagnostic position) and not a comorbidity (usually recorded in the third to the twentieth diagnostic position of an episode).

Continuous outcomes that will be defined in CPRD data, include HbA1c, SBP, DBP, HDL, LDL, total cholesterol, and triglycerides identified by searching the CPRD laboratory result data for recorded measures. eGFR will be defined by using serum creatinine (SCr) recorded in CPRD laboratory result data. The 2009 CKD-EPI equation, recommended by NICE,[29] will be used to calculate eGFR, without adjusting for ethnicity. BMI will be defined using a previously developed algorithm for CPRD data.[28] We will consider outcome values recorded within ±3 months from each follow-up time point of interest, to minimise outcome data missingness. The closest value to the follow-up time point of interest (e.g., 1 year) will be used, should there be >1 value recorded in the outcome time window.

Follow-up for all patients will be censored at the date of death, transfer out of the practice, end of data collection, or at 24-months follow-up, whichever is earliest. For outcomes measured in CPRD only (all continuous outcomes), the end of follow-up is 31 December 2021 (the end of primary care follow-up data). For outcomes measured in CPRD, HES,

and/or ONS, the end of follow-up is 31 March 2021 (the end of secondary care and ONS follow-up data).

3.6 Potential confounders/covariables

Using primary care data collected prior to second-line treatment initiation, we will extract information on patient characteristics (age, sex, ethnicity), time since T2DM diagnosis, time since first-line treatment initiation, and general practice size (reported as the number of patients registered with the GP in 2014). Using both primary and secondary care data, we will also identify relevant co-prescriptions prescribed within 60 days of baseline (renin-angiotensin system inhibitors (RASi), statins) and comorbidities at baseline (previous myocardial infarction (MI), unstable angina, stroke, hypoglycaemia, congestive heart failure (CHF), history of cancer (any), history of proteinuria, advanced eye disease, and lower extremity amputation). We will describe impaired kidney function defined using eGFR measures from primary care, with cut points defined by the Kidney Disease Improving Global Outcomes (KDIGO) guidelines for CKD, but without requiring two measures 3 months apart.[30] We will also describe the most recent measure of HbA1c, SBP, DBP, eGFR and BMI,[28] prior to baseline, as well as smoking and alcohol status using primary care records. IMD quintiles will be used to describe deprivation status. All codelists used to derive covariables will be published alongside any papers.

Table 4: List of covariates and the data sources used to define them

Covariable	Data source	Details
Age	CPRD	CPRD Age at baseline derived using the year of birth

Covariable	Data source	Details
Sex	CPRD	CPRD Sex recorded in CPRD
Ethnicity	CPRD, HES	CPRD Clinical code (Read or Snomed) indicating ethnicity, further categorised into four categories (White, South Asian, Black, Mixed/Other) HES Demographic data entered at in-patient hospitalisation, further categorised into four categories (White, South Asian, Black, Mixed/Other) Where CPRD ethnicity is missing, HES ethnicity is used to define people's ethnicity. Where ethnicities disagree, that recorded in CPRD is used.
Time since type 2 diabetes diagnosis	CPRD	CPRD Days between the first diagnosis code (Read or Snomed) for T2DM and baseline
Time on first-line (metformin monotherapy)	CPRD	CPRD Days between the first prescription for metformin and baseline
GP size	CPRD	CPRD Number of patients actively registered with the GP to which the patient belongs, derived using the CPRD denominator file, uses 2014 figures
NHS Region	CPRD	CPRD The region in which the GP practice is located to which each patient is registered. Regions include: East of England, London, Midlands, North East and Yorkshire, North West, South East, and South West
Co-prescriptions prescribed within 60 days of baseline (including RASi and statins)	CPRD	CPRD At least one prescription for the drug class of interest in the prescription history in the primary care record, within 60 days of baseline.
Comorbidities at baseline defined in primary and secondary care (including previous MI, unstable angina, stroke,	CPRD and HES	CPRD Diagnosis code (Read or Snomed) for each comorbidity prior to or the same day as baseline HES

Covariable	Data source	Details
hypoglycaemia, CHF)		Diagnosis code (ICD-10) for each comorbidity prior to or the same day as baseline in any diagnostic position of any episode of a spell
Comorbidities at baseline defined in primary care (cancer (any), advanced eye disease, lower extremity amputation, proteinuria)	CPRD	CPRD Diagnosis code (Read or Snomed) for comorbidity prior to or the same day as baseline
HbA1c	CPRD	CPRD Laboratory test recording the most recent HbA1c recorded within 180 days prior to baseline. Units reported as mmol/mol (tests recording HbA1c in % will be converted to mmol/mol).
eGFR and eGFR/CKD status	CPRD	CPRD Using the eGFR derived from serum creatinine using the CKD-EPI equation without adjustment for ethnicity recorded within 540 days prior to baseline, we will group people as either having eGFR≥60mL/min/1.73m^2 or eGFR<60mL/min/1.73m^2 (indicating impaired kidney function)
SBP and DBP	CPRD	CPRD Clinical measures captured in CPRD within 540 days prior to baseline
BMI	CPRD	CPRD BMI derived from weight and height measures entered by the GP (preferred), or BMI entered directly by the GP
Smoking status	CPRD	CPRD Clinical codes describing smoking status in the primary care record, using an algorithm previously defined in CPRD data
Alcohol status	CPRD	CPRD Clinical codes describing alcohol intake in the primary care record, using an algorithm previously defined in CPRD data
In-patient hospitalisation (any reason) in the past year	HES	HES At least one spell (hospitalisation) recorded in the patient's secondary care record (HES admitted patient care record) in the year prior to baseline

4. Analysis

Observational studies that adjust only for measured confounding variables are liable to report biased estimates of treatment effectiveness. We propose an IV study design as this can provide accurate estimates of effectiveness, even when there are unmeasured differences between the comparison groups.[5] The proposed IV exploits the wide variation across NHS CCGs (now ICSs) in the proportion of people with T2DM prescribed either SU, DPP4is, or SGLT2is, in addition to metformin. We refer to this as the tendency to prescribe (TTP) a particular treatment. The IV will be the prescribing history in each CCG. This will be defined as the proportion of relevant prescriptions within the CCG for each second-line oral antidiabetic treatment during the 12 months preceding the second-line treatment initiation of the patient currently under consideration. The IV for the prescription of each second-line antidiabetic treatment is the TTP of the CCG rather than the individual GP, because the choice of second-line treatment may involve the hospital diabetologist, the GP, other health-care professionals, and the patient.

4.1 The key IV assumptions

For the CCG prescribing history to be a valid instrument for the treatment prescribed, it must:

i) Strongly predict the treatment prescribed;

ii) Be independent of unmeasured baseline covariates; and

iii) Affect the outcome only through the treatment prescribed.[5]

The IV design will lead to bias if the prescribing history of the CCG has a direct effect on the outcome. For instance, a valid IV encourages treatment receipt, in this case of alternative

drugs, but does not have a direct effect on the outcome, for example HbA1c, except through the treatment received (see Figure 1). Our study draws from literature which shows that clinicians' prescribing history can be an IV when it strongly predicts the treatment offered, but does not have an independent effect on the outcome.[3,15,31] Our proposed IV design follows research in pharmaco-epidemiology that uses provider preference as an instrument for treatment prescribed.[15]

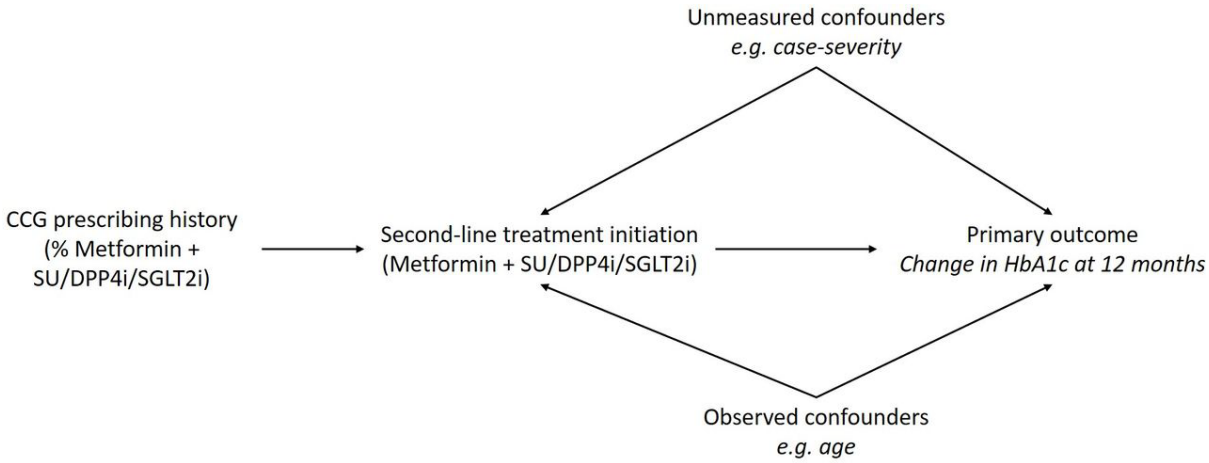

Figure 1: IV design to be applied in the PERMIT study[1]

Clinicians' prescribing history is likely to be independent of unmeasured confounders, since most patients choose to attend their local GP practice without considering their prescribing history. It seems unlikely that the CCGs prescribing history would have a direct effect on the patients' outcomes. One potential concern is that those time periods or CCGs which have a higher TTP the more recently available drug class of SGLT2i, might be associated with better quality of care provided. We address this concern by including covariates for 'time period' and NHS region within the IV analysis (see later section).

4.2 Checking IV assumptions

We will assess whether TTP meets the criteria for a valid and strong IV for each treatment of interest. First, we will assess whether the CCG-level TTP for each treatment is strongly associated with prescription of that treatment by reporting the Cragg-Donald F-statistic. Secondly, while it is not possible to assess the assumption that the IV is uncorrelated with the outcomes, except through the intervention, we will assess whether the CCG-level TTP balances the observed covariates.[32,33]

4.3 Accounting for defiers/always takers

Most IV applications assume that there are 'no defiers' also known as the monotonicity assumption. This means that a patient's treatment status is influenced by the status of the instrument (i.e., a patient is more likely to receive the treatment their CCG tends to prescribe more often than any of the other two alternatives). It is also assumed there are no 'always takers' or 'never takers' implying that there are no patients who would definitely receive any one of the three classes of treatment irrespective of their CCG's TTP.[34] That is, we assume the probability for receiving any of the three second-line treatments is between 0 and 1 for all people included in the study.

4.4 Two stage Least Squares (2SLS)

A common approach to IV analysis is to undertake two-stage least squares (2SLS) estimation. We will undertake preliminary analyses with 2SLS, whereby in the first stage, we will estimate a linear probability model for each of the treatments of interest (DPP4i or SGLT2i using SU as the reference category) on the instruments (TTP for either treatment), time fixed effects, and the covariates (described above). In the second stage, we will regress the outcome of interest on the covariates and the predicted propensity for each treatment

obtained from the first stage. The two stages will be estimated jointly so that standard errors reflect the uncertainty of both stages.

While 2SLS will be a useful preliminary model to estimate the relative treatment effects, this approach has a major limitation as it requires highly implausible assumptions to estimate average treatment effects across the whole population of interest. Hence, the main analysis will use an approach, two-stage residual inclusion (2SRI),[35] that can provide estimates that apply across the full population and sub-populations of interest.

4.5 Two stage Residual Inclusion (2SRI)

The two-stage residual inclusion (2SRI) is an IV approach that relies on concepts underlying control function methods to address the risk of bias from unmeasured confounding.[35] This approach includes the generalised residuals from the first-stage probit model, in the second-stage outcome models.[35] Unlike 2SLS, the 2SRI approach, when applied to a binary or categorical treatment such as the three-way comparison in the PERMIT study, aims to estimate the Average Treatment Effects (ATEs) for the overall population and subpopulations of interest.

In the second stage outcome models, we will use ordinary least squares models for continuous outcomes (e.g., 12 months HbA1c), and Cox proportional hazards models that account for individual frailty[36] for censored outcomes (e.g. time to 3-point MACE). The second-stage outcome models for both the continuous and time to event measures, will include the generalised residuals from the first stage, all measured baseline covariates, NHS region and time period. For the censored outcomes, the concern that the underlying proportional hazards (PH) assumption may not hold, due to changes in the hazard ratios comparing second-line antidiabetic treatments over time will be considered graphically and

with appropriate tests. If there is evidence that the PH assumption is violated then estimates of the relative treatment effect will be reported by including a time period by treatment interaction term in the respective models. All standard errors will be calculated with non-parametric bootstrapping, and 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. From these models, we will estimate the relative effect of the prescription of DPP4i versus SU, SGLT2i vs SU and SGLT2i vs DPP4i on outcomes across all the people included in the study.

In addition to the covariates included in Table 4, the methods described in sections 4.4 and 4.5 will also consider the quadratic forms of age and baseline HbA1c as well as two sets of interactions. The first set of interactions are those between baseline HbA1c with age, sex and baseline BMI. The second set of interactions are the products of the IV (for the first stage models) or the treatment indicator variables (for the second stage models) with baseline HbA1c, eGFR, BMI, systolic blood pressure and age.

To prevent overfitting the 2SLS and 2SRI models, we will use the lasso Regression algorithm to inform which of the interactions above are relevant in each case. The lasso aims to find the set of coefficients that minimise the sum-of-squares loss function subject to a constraint on the sum of absolute values of coefficients.[37] This results in a linear regression in which only a small number of covariates have non-zero coefficient that can then be included in the model in question. In particular, we will use the rigorous lasso data-driven method that guarantees that the covariates in Table 4 are always included in the models and will only penalise (or discard) variables in the interaction sets.[38]

To select the variables included in the estimation of the 2SLS and 2SRI models, we will run the rigorous lasso for the first and second stage models for each outcome. The final set of covariates for each outcome will include all the covariates in **Table 4** plus the interactions that will be selected in at least one model of the respective outcome.

4.6 Subgroup analyses

To limit the number of comparisons in the main analysis we will focus on the subgroup of prime interest, which is people with versus without established CVD (prevalent heart failure, ischemic heart disease (IHD), unstable angina, or previous stroke, myocardial infarction). This subgroup was selected in line with hypotheses that suggests SGLT2i may have a differential effect on subsequent rates of adverse cardiovascular events versus DPP4i and SU, for people with T2DM with pre-existing CVD.[6,39]

In subsequent exploratory analyses we will also consider relative effectiveness according to the following additional subgroups:

- Age (younger than 50 years old, 50-59 years, 60-69 years and 70 years and older)
- Sex (male, female)
- Ethnicity (White, South Asian, Black, and mixed/other, and missing)
- BMI category (under/healthy weight (<25), overweight (25-29.9), obese I (30-34.9), obese II (35+ kg/m^2))
- Deprivation (index of multiple deprivation quintiles)
- Electronic frailty Index (non-frail (eFI 0-0.12), mild frailty (0.12–0.24), moderate frailty (0.24–0.36) and severe frailty (>0.36))[40]
- With/without heart failure

- With/without reduced eGFR (eGFR≥60 ml/min/1.73m^2 versus eGFR≥30 and <60 ml/min/1.73m^2)

In further analyses, we will explore alternative data-driven methods to identify other subgroups for whom each treatment is beneficial/harmful, for example people with different combinations of long-term conditions. We will interpret the exploratory and further subgroup analyses as 'hypothesis generating'.

4.7 Handling missing data, censoring and loss to follow-up

The PERMIT study uses routine linked data which raises several challenges for the statistical analysis. Missing data is present in continuous outcomes and, also, covariates which are used in the analysis of the continuous and TTE outcomes. For example, HbA1c information may not be available at the 12- month timepoint, because the patient does not attend the GP, within the requisite time period (between month 9 and month 15). Information on baseline covariates such as HbA1c prior to second-line treatment may also not be available within the requisite time window, and information for time-constant measures such as ethnicity, may not be available from either the CPRD or linked HES data. In addition to missing data, all patients are not fully followed from baseline to 5 years. For example, a patient enrolled in December 2020 will have 12 months follow-up to 2021 and can only be included in the analysis models for the continuous outcomes for the periods between baseline and 6 months or 1 year, as they are unobserved for subsequent timepoints.

Background to the handling of missing data, censoring and loss to follow-up in this study

The longitudinal follow-up of patients in the data is complicated by the presence of missing data, censoring and loss to follow-up. While a patient may be included in the study and alive

for the duration of follow-up, they may not have been seen in primary care and will thus have unobserved values between baseline and the end of study follow-up which we term `intermittent missingness'. We propose using multiple imputation (MI)[41] to handle both missing values in the categorical or continuous baseline covariates (**Table 4**), and intermittent missingness in the continuous outcomes. The rationale for using MI is to impute these missing values with plausible substitutes based on the distribution of the observed data. Due to the non-linear trajectory of the continuous outcomes, it is best to utilise all information when imputing the outcome values. MI has the advantage of accounting for uncertainty in the imputed value while also incorporating observed relationships between the variable being imputed and other variables in the dataset. Therefore, for an unobserved outcome at time *t* we are able to incorporate the outcome values at all other times. For example, a patient's observed HbA1c values from baseline, 6 months and 2-5 years would be used to impute their unobserved year 1 HbA1c value, in addition to any auxiliary information which would improve the imputed value.

For patients who do not have full follow-up from baseline to 5 years due to death or censoring, it is complicated to include all follow-up measurements within any imputation process, and the nature of the form of end of follow-up will be recognised. A patient may be censored in the analysis models due to (i) end of study follow-up or (ii) the patient or GP practice no longer contributing to CPRD. We assume that these patients are censored 'completely at random'. This assumption would seem plausible as this censoring pertains to administrative reasons or due to the end of the follow-up period, which are unlikely to be related to the patient's characteristics of interest in our analyses, such as their prognosis. A patient who died or was censored at time point *t,* where $t = 0.5, 1, 2, \ldots, 5$ years, will be included in the continuous outcome analysis models described in sections 4.4-4.6 and any

missing information in covariates or outcomes from baseline to time *t* will need to be handled. For people who have died, their corresponding missing values prior to death are likely to differ to those who were alive at a given follow-up timepoint, and hence missing values for these people will be imputed separately from those with full follow-up, or those censored due to reasons other than death. While the implementation of the imputation models may for simplicity allow values to be imputed for timepoints after the time of censoring or death, we will respect censoring rules by dropping these imputed values prior to analysis.

A person who died or whose data are censored may be lost to follow-up prior to the time of death or censoring. For example, a patient may have observed outcomes at timepoints prior to year 2, and is then lost to follow-up (i.e. no observed values for any outcomes of interest from time *t* onwards) and does not have measurements at years 3 and 4, before being censored at year 5. The unobserved values from year 3 to 5 (monotone missingness) will be imputed.

Imputation steps

Missing data will therefore be addressed using multiple imputation (MI)[41] in the main analysis while respecting the reasons for any censoring (death or other reasons).[42] The MI method used will be MI by chained equations (MICE)[43] which will be used to generate 5 imputed datasets (M=5). All variables with missing data (Covariates: Index of multiple deprivation, baseline HbA1c, baseline eGFR, baseline BMI, Smoking status and Alcohol status, ethnicity; Continuous outcomes at timepoint *t*: HbA1c, eGFR, BMI, SBP) will be imputed by predictive mean matching with 10 donors[44] assuming that these data are

'missing at random' (MAR). This assumption implies that this intermittent missingness is at random, i.e., at random conditional on all other measures in the model including all preceding and subsequent levels of the measure in question, and the levels of any measures that were available at the timepoint in question. Some measurements taken repeatedly over time, e.g., HbA1c and BMI, will be missing at baseline for some individuals, and the same rationale for supporting the underlying MAR assumption would apply here as for outcomes with intermittent missingness, given that measurements for time periods prior to baseline and during the subsequent follow-up periods will be available for the imputation models.

For the time-constant baseline measures, the covariate with the greatest proportion of missing values is ethnicity. Previous literature has shown that conducting a MAR analysis for ethnicity can lead to similar point estimates as implementing missing data methods under the missing not at random assumption.[45,46] Here, our base case analysis will use multiple imputation for ethnicity, along with the other covariates, and we will examine robustness to the assumed missing data mechanism by undertaking complete-case analysis in a sensitivity analysis.

Squared terms for baseline continuous covariates (HbA1c, Age) and interaction terms will be treated as `just another variable' (JAV)[47] and imputed using PMM similarly to other baseline covariates.

The imputation model will be stratified by the assignment of second-line treatment (DPP4i, SU, SGLT2i), and status across the follow-up period (full follow-up; death; censored due to end of follow-up or patient/GP stopped contributing to CPRD) which will lead to a total of nine imputation models. The continuous outcome imputation models will include (i) all

covariates from the analysis model and (ii) the continuous outcome variables for all time points. The baseline covariate imputation models will adjust for the same covariates as the continuous outcome models ((i) and (ii)), while also including time-to-event information[48] (the event indicators and Nelson-Aalen estimates for each TTE outcome). This will ensure congeniality between the imputed baseline values and both the continuous and TTE outcomes. The same imputed datasets will be used to investigate an interaction between existing CVD and second line treatment, by relying on PMM to relax model assumptions.

Imputing individuals who died

For individuals who died during the follow-up period, the imputation models for baseline covariates and continuous outcomes will include a continuous variable for the time to death from baseline, to recognise that this may be predictive of the missing outcome. The imputation process will impute values for the continuous outcomes at time *t* where *t* is a time post-death. However, for the subsequent outcome analysis models, all imputed values after the time of death will be discarded.

Imputing individuals who are censored

The reasons for censoring are (i) the end of the data extraction period (i.e., the end of follow-up) and (ii) the person or GP practice stop contributing to CPRD. Similarly, to those patients who are followed to the end of study, or those who have died, for those who are censored, one imputation model will be used (stratified by treatment) to impute missing covariate and continuous outcome values. Censoring rules will subsequently be applied at

the outcome analysis stage to remove any imputed continuous outcome values after the point of censoring.

Post-imputation procedures

The analysis model of interest, will be applied to each of the five imputed datasets. The relative treatment effects will be estimated using two-stage residual inclusion IV (with a frailty inclusion[36] for time-to-event outcomes when using Cox proportional hazards). Rubin's rules[49] will be applied to obtain an overall treatment effect:

$$\hat{\bar{\theta}}_d = M^{-1} \sum_{m=1}^{M} \hat{\theta}_{m,d}$$

where *d* = SGLT2i or DPP4i.

Confidence intervals for the treatment effects will be estimated using bootstrap sampling (BS). We will first bootstrap sample and then apply MI within each sample (BS-then-MI[45,46]). The original unimputed dataset will be used to draw B=500 bootstrap samples (this may be adjusted depending on Monte Carlo error), with the bootstrap draw stratified by GP region, treatment group, and follow-up status (fully-followed, death, censored) to maintain similar sampling patterns within each bootstrap sample.

Within each bootstrap sample $b = 1, \ldots, B$, we will take the same approach to handling missing data and implementing the analysis model. Rubin's rules will be applied to the *M* imputed datasets of bootstrap sample *b* to get an overall treatment effect for drug *d*:

$$\hat{\bar{\theta}}_{b,d} = M^{-1} \sum_{m=1}^{M} \hat{\theta}_{m,d}$$

The 500 estimates will be used to estimate variance and calculate *t*-based confidence intervals.

R[50] will be used for (i) the imputation of missing values in this analysis (MICE package[51]) and (ii) the analysis of time-to-event outcomes using individual frailty.[52] Stata 17[53] will be used for the analysis of continuous outcomes.

4.8 Alternative analyses

To investigate the extent to which the findings from the base case analysis are robust to alternative assumptions, we will undertake sensitivity analyses, including those described in **Table 5.**

Table 5: Proposed sensitivity analyses

Type of analysis alternative	**Cohort (imputed or complete case)**	**Rationale**
Base case	Imputed	
Alternative analysis – complete case	Complete cases	Assess whether the results are sensitive to an alternative assumption about the missing data.
Alternative analysis – 2SLS IV model	Complete case	Compare the results of the base case (2SRI IV analysis) with those from an alternative IV model (2SLS).
Alternative analysis – multivariable regression adjusted for measured confounders only	Complete case	Determine to what extent unmeasured confounding may be impacting the results in the base case.
Alternative analysis – extend follow-up to 5 years rather than 2 years	Complete cases	For the main analyses, we will consider outcomes up to 2-years follow-up, assuming that beyond this time point most people will have switched antidiabetic treatment. We will extend follow-up to 5 years in this alternative analysis, with

Type of analysis alternative	Cohort (imputed or complete case)	Rationale
		careful interpretation on treatment effects beyond the 2-year time point.

Appendix

Amendments to the originally published PERMIT protocol[1] (Bidulka et al., 2021)

Link to published protocol paper:
https://bmjopen.bmj.com/content/11/9/e046912.abstract

Original text from the protocol is written in column 2, and amended text is written in column 3. The pieces of text that are bolded are the specific aspects of the original protocol which are amended.

Amendment number	Original plan (as per Bidulka et al. 2021)	Amended plan
1	Study population: The study population will include people registered with **a CPRD-contributing practice,** aged 18 years or older, diagnosed with T2DM who intensify antidiabetic treatment from metformin-monotherapy to a combination of metformin and SU, DPP4i or SGLT2i (second-line treatment) between **2014 and 2020.**	Study population: The study population will include people registered with **a CPRD-contributing practice with HES/ONS/IMD linkage available,** aged 18 years or older, diagnosed with T2DM who intensify antidiabetic treatment from metformin-monotherapy to a combination of metformin and SU, DPP4i or SGLT2i (second-line treatment) between **1 January 2015 and 31 December 2020.** Justification: The time to event outcomes require HES/ONS/IMD linkage. Further, additional data from HES (e.g., ethnicity data, comorbidity data), ONS (gold-standard death date in England and Wales), and the IMD (small-area deprivation) are important to reduce data missingness and misclassification. When designing the study, we anticipated that only 70% of the CPRD population in England would be eligible for HES-linkage based on data resource profiles for CPRD.[16,17] However, in our updated dataset the proportion eligible for linkage is 91%. Thus, we feel that the balance between maximising study power

Amendment number	Original plan (as per Bidulka et al. 2021)	Amended plan
		for all outcomes versus the benefits in analysing and interpreting results for one clearly defined cohort is in favour of latter. In addition, we have changed the study time period in which people can enter the study. We require one year of historical prescribing data from each person's index date to define the instrument in our instrumental variable analysis – thus we must exclude people who initiate second-line antidiabetic treatment in 2014, since prior to this year SGLT2i were not widely prescribed. We also extended the dataset to allow for the inclusion of people who intensify second-line antidiabetic treatment up until 31 December 2020.
2	End of follow-up: People will remain exposed until the date the data are censored by death, the patient leaving the GP practice, the GP practice stops contributing to CPRD, **or 31 July 2020.**	End of follow-up: People will remain exposed until the date the data are censored by death, the patient leaving the GP practice, the GP practice stops contributing to CPRD, **or 31 December 2021 (for outcomes defined in primary care only) and 31 March 2021 (for outcomes defined in secondary care (HES) or ONS data).** Justification: The dataset was extended to allow for follow-up until 31 December 2021 for outcomes defined in primary care data. For outcomes defined in secondary care data, we must end follow-up at 31 March 2021 as this was the maximum amount of follow-up available at the time of study data extraction.
3	Covariates:	Covariates:

Amendment number	Original plan (as per Bidulka et al. 2021)	Amended plan
	We will also identify HbA1c, systolic blood pressure (SBP), diastolic blood pressure (DBP), eGFR, body weight, and body-mass index (BMI) using values recorded **in the 180 days period before the second-line antidiabetic treatment initiation date** in the primary care record.	We will also identify HbA1c, systolic blood pressure (SBP), diastolic blood pressure (DBP), eGFR, body weight, and body-mass index (BMI) using values recorded in the 180 days (HbA1c) and 540 days (SBP, DBP, eGFR, BMI) period before the second-line antidiabetic treatment initiation date in the primary care record. Justification: We follow precedent research that used a 540 day window pre-baseline to define these clinical measures at baseline.[25] However, we still require HbA1c to have been measured within 180 days as we expect older values of HbA1c to be unrepresentative of the patient's HbA1c status at baseline.
4	Primary outcome: The primary outcome for objective 2 (short-term relative effectiveness) will be absolute change in HbA1c**%** at 12 months follow-up.	Primary outcome: The primary outcome for objective 2 (short-term relative effectiveness) will be absolute change in HbA1c (**mmol/mol**) at 12 months follow-up. Justification: UK is aligning with Europe and reporting HbA1c using International Federation of Clinical Chemistry (IFCC) units (mmol/mol) rather than Diabetes Control and Complications Trial (DCCT) unit (%).
5	Secondary outcomes: We will also report change in HbA1c at 6–18, 24–30 and 36months follow-up, again using the closest HbA1c measure in the 3 months before and after the follow-up time point of interest.	Secondary outcomes: We will also report change in HbA1c**, eGFR, SBP, and BMI** at **6-, 24-, 36-, 48-, and 60-months follow-up,** again using the closest outcome measure in the 3 months before and after the follow-up time point of interest. Justification:

Amendment number	Original plan (as per Bidulka et al. 2021)	Amended plan
		Our dataset was updated to include a maximum follow-up time of 7 years. We therefore increased the duration of follow-up time points of interest, and simplified to be at yearly intervals (other than the first 6-month time-point). While we planned to investigate eGFR, BMI, and SBP at 12 months follow-up, we also plan to investigate these outcomes at every other follow-up time point of interest as per HbA1c.
6	Secondary outcomes: Outcomes for long-term relative effectiveness (objective 3) will include macrovascular and microvascular conditions **such as CV outcomes** (MI, CHF, unstable angina, stroke)**, renal outcomes (**nephropathy, ESRD, 40% decline in eGFR from baseline**)** and lower limb amputation.	Secondary outcomes: Outcomes for long-term relative effectiveness (objective 3) will include macrovascular and microvascular conditions **such as CV outcomes (3-point major adverse cardiovascular event (MACE), a composite outcome of myocardial infarction, stroke, and all-cause mortality),** MI, CHF, unstable angina, stroke**), renal outcomes (a composite kidney outcome (40% decline in eGFR from baseline, end-stage kidney/renal disease (ESKD), and all-cause mortality), as well as** nephropathy, ESRD, 40% decline in eGFR from baseline**)** and lower limb amputation. Justification: We seek to emulate trials which compare these second-line antidiabetic drugs, which often report MACE and a composite kidney outcome.[8,54] The components of these outcomes were already specified in the protocol, and pre-specified adding the composite end-points before conducting analyses.
7	Secondary outcomes: A 40% decline in eGFR will be	Secondary outcomes:

Amendment number	Original plan (as per Bidulka et al. 2021)	Amended plan
	defined as an eGFR measure **≤40% of baseline eGFR.**	A 40% decline in eGFR will be defined as an eGFR measure **≤60% of baseline eGFR.** Justification: This was a mistake in the original protocol – the 40% decline should represent an eGFR measure that is ≤60% of baseline eGFR.
8	Analytical approach (Objective 1): We will describe trends in prescribing for T2DM second-line treatment for the duration of the study period across the UK and between CCGs. This analysis will update previous research which described the same second-line treatment use in the UK from 2000 to 2017, and will employ similar methods.	Analytical approach (Objective 1): We will describe trends in prescribing for T2DM second-line treatment for the duration of the study period across the UK and between CCGs, **particularly with respect to clinically important factors predicting which type of second-line antidiabetic treatment people are prescribed, such as ethnicity and deprivation.** This analysis will update previous research which described the same second-line treatment use in the UK from 2000 to 2017, and will employ similar methods. Justification: Understanding factors which predict prescribing for particular second-line antidiabetic treatments is helpful to design the instrumental variable analysis. Previous work by Wilkinson et al (2018)[24] described factors associated with choice of second-line antidiabetic treatment. We build off this work in understanding whether there are sociodemographic disparities in which type of second-line antidiabetic treatment is prescribed.
9	Planned analyses – the IV: We will provide **personalised estimates of treatment**	Planned analyses – the IV: Due to challenges in developing the methodology to compare three

Amendment number	Original plan (as per Bidulka et al. 2021)	Amended plan
	effectiveness using the local IV (LIV) approach to predict the counterfactual outcomes that each person would experience if they were prescribed each second-line treatment.	rather than two treatments using the local IV approach, we will instead use the two-stage residual inclusions (2SRI) model to conduct this analysis. This approach also enables treatment effectiveness to be reported for the overall populations and subpopulations of prime interest.
10	Missing data: **In our primary analysis, we will use a complete-case approach based on the main potential confounders listed in the covariates section. We will conduct secondary analyses using complete cases for the full list of potential confounders, including those expected to have a high proportion of missing data (see covariates section), which we do not expect to be missing at random. Because we cannot assume covariate measurements are missing at random and the IV model is computationally intensive, we will not use multiple imputation.** We will adopt two main approaches based on the type of missingness for outcome data: (1) linear interpolation using values recorded during follow-up, and (2) inverse probability weighting (IPW) to those people lost to follow-up with no subsequent outcome measure.	Missing data: We now propose using multiple imputation (MI) to handle missing values in covariates and intermittent missingness in the continuous outcomes as it will impute unobserved values with plausible substitutes based on the distribution of the observed data. We will handle loss to follow-up with inverse probability weighting. Justification. The initial data descriptions highlighted the non-linear trajectory of the continuous outcomes, it is best to utilise all information when imputing the outcome values. MI has the advantage of accounting for uncertainty in the imputed value while also incorporating observed relationships between the variable being imputed and other variables in the dataset. For loss to follow-up, previous work has shown the problems of using MI for imputing for timepoints beyond the observed data, and that IPW that only relies on baseline values to reweight the observations is more appropriate for handling this problem.

6. References

1. 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(9):e046912. doi:10.1136/bmjopen-2020-046912
2. 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, Obesity and Metabolism*. 2023;25(1):282-292. doi:https://doi.org/10.1111/dom.14874
3. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758-764. doi:10.1093/aje/kwv254
4. Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. *Jama*. 2022;328(24):2446-2447. doi:10.1001/jama.2022.21383
5. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med*. Jun 15 2014;33(13):2297-340. doi:10.1002/sim.6128
6. NICE guideline [NG28]: Type 2 diabetes in adults: management. Web. NICE. Accessed 3 March, 2022. https://www.nice.org.uk/guidance/ng28/chapter/Recommendations#reviewing-drug-treatments
7. Buse JB, Wexler DJ, Tsapas A, et al. 2019 Update to: Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes care*. 2020;43(2):487. doi:10.2337/dci19-0066
8. Glycemia Reduction in Type 2 Diabetes — Microvascular and Cardiovascular Outcomes. *New England Journal of Medicine*. 2022/09/22 2022;387(12):1075-1088. doi:10.1056/NEJMoa2200436
9. Glycemia Reduction in Type 2 Diabetes — Glycemic Outcomes. *New England Journal of Medicine*. 2022/09/22 2022;387(12):1063-1074. doi:10.1056/NEJMoa2200433
10. Wexler DJ, de Boer IH, Ghosh A, et al. Comparative Effects of Glucose-Lowering Medications on Kidney Outcomes in Type 2 Diabetes: The GRADE Randomized Clinical Trial. *JAMA internal medicine*. 2023;doi:10.1001/jamainternmed.2023.1487
11. Type 2 diabetes [B] Pharmacological therapies with cardiovascular and other benefits in people with type 2 diabetes, NICE Guideline NG28. Web. National Institute for Health and Care Excellence (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
12. Rosenstock J, Kahn SE, Johansen OE, et al. Effect of Linagliptin vs Glimepiride on Major Adverse Cardiovascular Outcomes in Patients With Type 2 Diabetes: The CAROLINA Randomized Clinical Trial. *Jama*. Sep 19 2019;322(12):1155-66. doi:10.1001/jama.2019.13772
13. 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(7):e022768. doi:10.1136/bmjopen-2018-022768
14. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. Apr 15 2016;183(8):758-64. doi:10.1093/aje/kwv254

15. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *The international journal of biostatistics*. 2007;3(1):Article 14. doi:10.2202/1557-4679.1072
16. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. Jun 2015;44(3):827-36. doi:10.1093/ije/dyv098
17. Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *International Journal of Epidemiology*. 2019;48(6):1740-1740g. doi:10.1093/ije/dyz034
18. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *International journal of epidemiology*. 2017;46(4):1093-1093i. doi:10.1093/ije/dyx015
19. National statistics: English indices of deprivation 2019. Web. Ministry of Housing, Communities & Local Government. 2022. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019
20. Penney B. *The English Indices of Deprivation 2019 (IoD2019)*. 2019. Accessed 4 March 2022. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/835115/IoD2019_Statistical_Release.pdf
21. Guidance for industry: diabetes mellitus: developing drugs and therapeutic biologics for treatment and prevention. US Food and Drug Administration. https://www.regulations.gov/document/FDA-2008-D-0118-0003
22. Farmer AJ, Rodgers LR, Lonergan M, et al. Adherence to Oral Glucose-Lowering Therapies and Associations With 1-Year HbA1c: A Retrospective Cohort Analysis in a Large Primary Care Database. *Diabetes care*. Feb 2016;39(2):258-263. doi:10.2337/dc15-1194
23. Ranstam J. Hypothesis-generating and confirmatory studies, Bonferroni correction, and pre-specification of trial endpoints. *Acta Orthop*. Aug 2019;90(4):297. doi:10.1080/17453674.2019.1612624
24. 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;10:1639-1648. doi:10.2147/CLEP.S176142
25. 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, Obesity and Metabolism*. 2020;22(5):847-856. doi:https://doi.org/10.1111/dom.13970
26. 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. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. Dec 2014;64(6):821-35. doi:10.1053/j.ajkd.2014.07.030
27. Levin A, Agarwal R, Herrington WG, et al. International consensus definitions of clinical trial outcomes for kidney failure: 2020. *Kidney Int*. Oct 2020;98(4):849-859. doi:10.1016/j.kint.2020.07.013
28. 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(9):e003389. doi:10.1136/bmjopen-2013-003389

29. Levey AS, Stevens LA, Schmid CH, et al. A New Equation to Estimate Glomerular Filtration Rate. *Annals of Internal Medicine*. 2009/05/05 2009;150(9):604-612. doi:10.7326/0003-4819-150-9-200905050-00006
30. Stevens PE, Levin A. Evaluation and Management of Chronic Kidney Disease: Synopsis of the Kidney Disease: Improving Global Outcomes 2012 Clinical Practice Guideline. *Annals of Internal Medicine*. 2013/06/04 2013;158(11):825-830. doi:10.7326/0003-4819-158-11-201306040-00007
31. Widding-Havneraas T, Chaulagain A, Lyhmann I, et al. Preference-based instrumental variables in health research rely on important and underreported assumptions: a systematic review. *J Clin Epidemiol*. Nov 2021;139:269-278. doi:10.1016/j.jclinepi.2021.06.006
32. Basu A. ESTIMATING PERSON-CENTERED TREATMENT (PeT) EFFECTS USING INSTRUMENTAL VARIABLES: AN APPLICATION TO EVALUATING PROSTATE CANCER TREATMENTS. *J Appl Econ (Chichester Engl)*. June/July 2014;29(4):671-691. doi:10.1002/jae.2343
33. Basu A, Gore JL. Are Elderly Patients With Clinically Localized Prostate Cancer Overtreated? Exploring Heterogeneity in Survival Effects. *Medical care*. Jan 2015;53(1):79-86. doi:10.1097/mlr.0000000000000260
34. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol*. Dec 2009;62(12):1226-32. doi:10.1016/j.jclinepi.2008.12.005
35. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*. 2008/05/01/ 2008;27(3):531-543. doi:https://doi.org/10.1016/j.jhealeco.2007.09.009
36. Martínez-Camblor P, Mackenzie T, Staiger DO, Goodney PP, O'Malley AJ. Adjusting for bias introduced by instrumental variable estimation in the Cox proportional hazards model. *Biostatistics*. Jan 1 2019;20(1):80-96. doi:10.1093/biostatistics/kxx062
37. Kreif N, DiazOrdaz K. Machine Learning in Policy Evaluation: New Tools for Causal Inference. Oxford University Press; 2019.
38. Ahrens A, Hansen CB, Schaffer ME. lassopack: Model selection and prediction with regularized regression in Stata. *The Stata Journal*. 2020;20(1):176-235. doi:10.1177/1536867x20909697
39. Zheng SL, Roddick AJ, Aghar-Jaffar R, et al. Association Between Use of Sodium-Glucose Cotransporter 2 Inhibitors, Glucagon-like Peptide 1 Agonists, and Dipeptidyl Peptidase 4 Inhibitors With All-Cause Mortality in Patients With Type 2 Diabetes: A Systematic Review and Meta-analysis. *Jama*. 2018;319(15):1580-1591. doi:10.1001/jama.2018.3024
40. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age and Ageing*. 2016;45(3):353-360. doi:10.1093/ageing/afw039
41. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc.; 1987.
42. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*. 2013/06/01 2011;22(3):278-295. doi:10.1177/0962280210395740

43. van Buuren S, Oudshoorn, C.G.M. *Multivariate Imputation by Chained Equations: MICE V1.0 User's manual. TNO Report PG/VGZ/00.038*. 2000. http://www.multiple-imputation.com/
44. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology*. 2014/06/05 2014;14(1):75. doi:10.1186/1471-2288-14-75
45. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*. 2020/08/01 2020;584(7821):430-436. doi:10.1038/s41586-020-2521-4
46. Mathur R, Rentsch CT, Morton CE, et al. Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. *Lancet (London, England)*. May 8 2021;397(10286):1711-1724. doi:10.1016/s0140-6736(21)00634-6
47. Seaman SR, Bartlett JW, White IR. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC Medical Research Methodology*. 2012/04/10 2012;12(1):46. doi:10.1186/1471-2288-12-46
48. White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med*. Jul 10 2009;28(15):1982-98. doi:10.1002/sim.3618
49. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. Feb 20 2011;30(4):377-99. doi:10.1002/sim.4067
50. Team RC. R: A language and environment for statistical computing. Web. R Foundation for Statistical Computing. https://www.R-project.org/
51. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 12/12 2011;45(3):1 - 67. doi:10.18637/jss.v045.i03
52. Therneau T. A Package for Survival Analysis in R: R package version 3.4-0. https://CRAN.R-project.org/package=survival
53. StataCorp. Stata Statistical Software: Release 17. StataCorp LLC.
54. Empagliflozin in Patients with Chronic Kidney Disease. *New England Journal of Medicine*. 2022;388(2):117-127. doi:10.1056/NEJMoa2204233