Statistical Analysis Plan – The Trimaster Study

TRIAL FULL TITLE	TriMaster: Randomised Double-Blind Crossover study of a DPP4 inhibitor, SGLT2 inhibitor and thiazolidinedione as third line therapy in patients with type 2 diabetes who have suboptimal glycaemic control on dual therapy with metformin and a sulphonylurea
TRIAL REGISTRATION NUMBER	ISRCTN 12039221 ClinicalTrials number (NCT02653209) EudraCT number (2015-002790-38)
SAP VERSION	V9 - Final signed off
SAP VERSION DATE	11th March 2021
TRIAL STATISTICIAN	Dr Beverley Shields
TRIAL CHIEF INVESTIGATOR	Prof Andrew Hattersley
SAP AUTHOR	Dr Beverley Shields

SAP Signatures

I give my approval for the attached SAP entitled “TriMaster: Randomised Double-Blind Crossover study of a DPP4 inhibitor, SGLT2 inhibitor and thiazolidinedione as third line therapy in patients with type 2 diabetes who have suboptimal glycaemic control on dual therapy with metformin and a sulphonylurea” dated 11th March 2021

Chief Investigator

Name: Andrew Hattersley

Signature: 

Date: 18/03/2021

Statistician

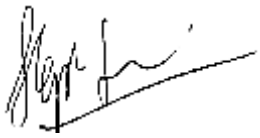
Name: Beverley Shields

Signature: 

Date: 11 March 2021

Independent Statistician

Name: Stephen Senn

Signature: 

Date: 17 March 2021

1. Background and Rationale:

The Trimaster trial is part of MASTERMIND, a MRC-funded study aiming to develop a stratified approach in Type 2 diabetes that will result in more effective use of glucose lowering therapy. Patients with Type 2 diabetes vary greatly in how well they respond to different diabetes drugs and whether they develop side effects to particular medications.

The rationale for a stratified approach is based on the following:

- Patients with Type 2 diabetes show considerable inter-individual variation in their underlying pathophysiology.
- The different classes of glucose-lowering therapies work by very different mechanisms of action.
- Pilot studies have shown that variation in response to therapy is, in part, robustly explained by differences in patients' underlying pathophysiology.

Type 2 diabetes is common (approx. 4% of the population) and most prescribing of relatively inexpensive therapy is in primary care. Therefore, identification of subgroups of patients who will respond to a given therapy needs to be based on clinical characteristics and readily available biomarkers in routine clinical care.

In this study, the research team aims to identify subgroups of patients that respond well or poorly to third-line therapies based on particular clinical characteristics such as BMI and renal function.

At the time of study start (November 2016), NICE guidelines recommended three potential oral therapies that could be used in addition to metformin and sulphonylureas, for treating suboptimal glycaemic control (HbA1c >=58mmol/mol): DPP4 inhibitors (DPP4i), SGLT2-inhibitors (SGLT2i) or thiazolidinediones (TZD). Preliminary data from the original MASTERMIND programme suggest obese patients are likely to respond better to TZDs[1] and non-obese patients are likely respond better to DPP4i[2] (from UK primary care data, Clinical Practice Research Datalink (CPRD)), and that patients with good renal function are likely to respond better to SGLT2i and patients with poor renal function are likely to respond better to DPP4i (industry trial data shared by Janssen).

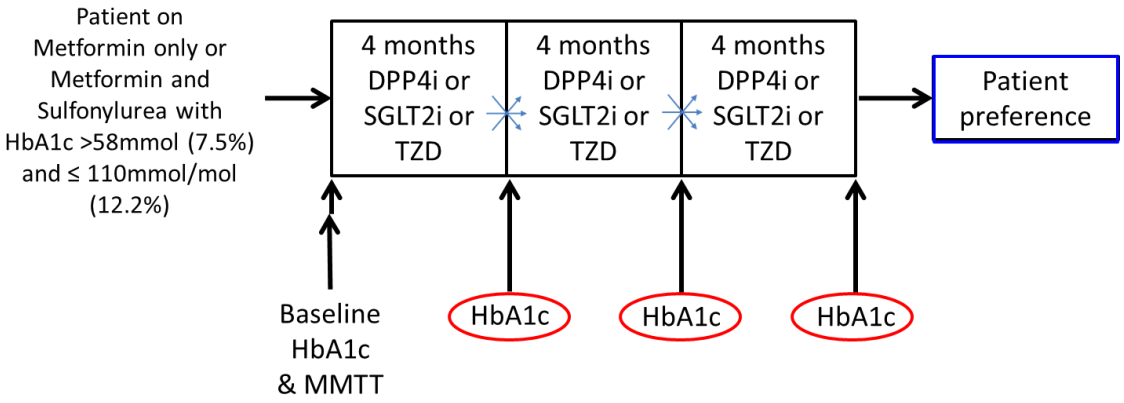
2. Purposes of analyses:

2.1 Study Objectives and Endpoints

The primary objective is to test hypothesised stratification using a three-period, three-treatment randomised double-blind crossover study of second/third line oral therapy in Type 2 diabetes, comparing TZD, DPP4i and SGLT2i (Figure 1). The primary outcome is achieved glycaemic control after 4 months on therapy (HbA1c measured at the end of each of the 4 month treatment periods) and secondary outcomes are patient preference, tolerability, and side effects.

The study aims to test two pre-specified hypotheses of drug response stratification based on drug mechanism of action and pharmacokinetics that are supported by routine clinical and trial data. All patients will receive all three drugs, but the analysis for the two separate hypotheses will each compare only two drugs at a time.

Figure 1: Diagram showing the study design for Trimaster



2.2 Hypotheses:

1. Patients with insulin resistance, characterised clinically by a raised BMI (>30kg/m^2), compared to non-obese patients, will:
   Respond well to pioglitazone, a TZD that works as an insulin sensitiser, **but less well to sitagliptin**, a DPP4i, which works through stimulating endogenous insulin secretion post-prandially.
2. Patients with modestly reduced estimated glomerular filtration rate (eGFR 60-90 mls/min/1.73m^2), compared to those with eGFR>90mls/min/1.73m^2, will:
   Respond less well to canagliflozin, a SGLT2 inhibitor, which works through inhibiting the active reabsorption of glucose in the proximal tubule, as the reduced eGFR will decrease the glucose-lowering efficacy, **but respond well to sitagliptin**, a DPP4i that is renally cleared, as the reduced eGFR will increase plasma DPP4i concentrations.

2.3 Primary objective

The primary objective of the trial is to test the above hypotheses, specifically:

a) To determine whether the difference in glycaemic response to pioglitazone and sitagliptin is different in obese (BMI>30kg/m^2) and non-obese patients (BMI≤30kg/m^2).
b) To determine whether the difference in glycaemic response to sitagliptin and canagliflozin is different in patients with eGFR>90mls/min/1.73m^2 compared with eGFR 60-90mls/min/1.73m^2.

2.4 Secondary objectives:

The secondary objectives of the trial are to assess:

a) Tolerability (continuation/discontinuation of therapy), overall for each drug and comparing between hypothesised strata
b) Prevalence of side effects, overall for each drug and comparing between hypothesised strata, particularly known side effects of specific drugs to include: weight gain, hypoglycaemia, oedema, genital tract infection
c) Patient treatment preference, overall and also comparing between hypothesised strata (obese v non-obese and eGFR 60-90 v eGFR>90).

2.5 Primary Endpoint (see Section 14 for full definitions)**:**

The primary outcome of this crossover study is HbA1c measured at the end of each of the 4 month treatment periods.

2.6 Secondary Endpoints (see analysis section (section 14) for full definitions):

Tolerability: Whether patients tolerate (complete the full treatment period) or don't tolerate (stop their medication early) a given treatment will be recorded.

Side effects: Frequency and severity of recorded common side effects of all 3 medications in the trial will be recorded at the end of each treatment period.

Patient preference: Patients will be asked which treatment(s) they would prefer to take long term and the reason for their preference.

3 General Study Design and Plan

The study is a phase 4, randomised, double-blind, 3 way crossover study of a DPP4-inhibitor (sitagliptin), SGLT2 inhibitor (canagliflozin), and thiazolidinedione (pioglitazone) as second or third line therapy in patients with Type 2 diabetes who have suboptimal glycaemic control on therapy with either metformin alone or metformin and a sulphonylurea.

Minimising carryover and period effects: We do not anticipate carryover effects and have designed the study to limit potential carryover effects as far as possible (see section 3.1).
Any period effect in the maximum 8 months between on treatment HbA1cs is likely to be minimal as mean progression is 1.0 mmol/mol/year (personal communication E Pearson, data from DARTs Tayside population data).

3.1 Rationale for Trial Design:

Crossover Study: A crossover approach was chosen rather than a parallel group clinical trial design as the research team seek to establish individual patient responses to the three drug classes being evaluated. This requires each patient receiving all three treatments, thereby acting as their own control.

Three-way crossover study: A three-way crossover design was chosen rather than two two-way crossover studies to address each of the two hypotheses because this is a more efficient, quicker and cost effective approach requiring fewer patients. In addition, the resulting bioresource will be of greater value for subsequent hypothesis generation as it will link responses to three rather than two therapies.

No washout period: There are no washout periods between drugs. In the MASTERMIND pilot study, it was found that re-establishing a stable baseline when each trial therapy was discontinued was problematic, as patients experienced high glucose levels in the washout period leading to high numbers of drop outs. To avoid this problem, the major comparison will be the difference between drugs in terms of glycaemia (HbA1c) achieved on stable therapy avoiding the need for washout periods between treatments as the change from baseline is not assessed (see section 14 for full definition).

4 month treatment periods: To minimise potential carry-over effects, the treatment period for each drug will be 4 months. HbA1c reflects glucose levels over the preceding 8-12 week period with the glucose levels closest to the sample being taken having the greatest contribution[3 4]. Therefore, glucose in the first month of treatment will have minimal effect. All three drugs have half-lives of between 7 and 14 hours, so

their effects should be negligible after a week. Potential carryover and period effects will be assessed and described as part of the final analysis (see section 15.1).

4 Randomisation and Blinding

This study is a randomised controlled double-blind three-way crossover trial.

4.1 Randomisation: This study includes 3 drugs, providing 6 potential order combinations: ABC, ACB, BAC, BCA, CAB, CBA. All participants receive the three therapies in random order, according to one of these six possible treatment orders. A randomisation list comprising 660 of these order combinations (block size 12) was created by the study statistician using the randomisation procedure in StatsDirect (randomisation seed recorded). The list size of 660 was determined by the initial sample size (prior to study start) of 600, plus an additional 10% (additional combinations to be ignored if not needed). The randomisation schedule was provided to the study database team who assigned the order combinations from the list to each of the study sites in blocks of 12 to ensure even distribution of treatment orders at each site (total number of allocations determined according to target recruitment number). The actual allocations assigned to each site are stored securely in the database hidden from all other users, and importantly the study team including researchers and the study statistician, to ensure they remain blinded to the final allocations.

4.2 Blinding: Drugs were blinded by over-encapsulation (by Tayside Pharmaceuticals). The medication was labelled with a unique identification number and supplied directly to each study site pharmacy. Drugs were supplied to each site in batches of 12 to ensure a supply of 4 bottles of each drug. The total amount supplied to each site was based on anticipated recruitment and shelf-life of the drugs. The list of medication IDs assigned to the drugs provided to each site was sent directly to the study database team.

4.3 Matching of blinded drugs to randomisation allocation:
The randomisation schedule and list of medication IDs assigned to each site is securely stored in the study database to allow matching of medication IDs to patient allocation. At recruitment, the database looks up the next available ID on the randomisation schedule for that site to determine the allocated treatment order for that patient. At each study visit for a given patient, the database identifies the appropriate medication bottle ID at that site matching the allocated drug for that visit.

4.4 Statistician blinding
Once all patient data has been collected and cleaning performed by the study team, the statistician will perform a blinded review of exported data and dummy run of primary analysis as a final check before data lock and final sign off to the Statistical Analysis Plan (see section 7). Due to the nature of the crossover trial, at this stage, there will be no way of identifying the drugs patients receive as data will only be available by treatment period, and each treatment period consists of a mix of patients on all 3 drugs. Any final amendments to the data or statistical analysis plan following this review will be clearly documented. Following data lock, there will be two stages of analysis: one with drug allocations provided but in coded format, and one with final drug allocations revealed (see section 17).

5 Sample Size

This trial aims to test whether patients in a particular strata respond differently to drug A and drug B (either TZD and DPP4i or SGLT2i and DPP4i dependent on the hypothesis) compared with patients not in the strata. The required sample size will be dependent on the SD of change in HbA1c, the likely impact of the strata, and the chosen statistical power and significance level.

The SD of change in HbA1c on two different therapies in a crossover trial setting is estimated as 8.7mmol/mol (crossover trial of metformin v repaglinide[5]).

Preliminary data from UK primary care data (Clinical Practice Research Datalink) showed obese patients respond better to TZDs and non-obese patients respond better to DPP4is, with an overall difference in response between obese and non-obese strata of 3.1mmol/mol (equivalent to 0.36SDs). Similarly, patients with an eGFR>90ml/min/1.73m² show a better response to canagliflozin (clinical trial data shared by Janssen) whereas patients with an eGFR 60-90ml/min/1.73m² have a better response to sitagliptin, with an overall difference in response between these two strata of 3.0mmol/mol (equivalent to 0.35SDs).

Using 90% power, alpha=0.05, to detect a difference of 0.35SDs we require 172 patients in each strata, so 344 in total. To allow for the possibility of unequal numbers in each strata, the sample size increases to 358 patients assuming a 60:40% split (In patients with Type 2 Diabetes in CPRD the split was 52:48% for both strata (i.e. obese/non-obese and eGFR 60-90/eGFR>90)). Allowances also need to be made for potential drop out/exclusions that would rule out participants from primary analysis: 1) To allow for a conservative withdrawal rate of 15% (which could include patients dropping out at any stage of the crossover trial), this increases the study sample size to 422. 2) To allow for participants being excluded from primary analysis due to discontinuing at least one of the study drugs within the first 8 weeks on the study drug (conservatively estimated at 19%), the sample size increases to 520. This sample size is conservative as allows for all the drop outs/discontinuation to occur in the DPP4i arm, and thus affecting analyses for both hypotheses. Drop outs/exclusions of the other two study drugs would only impact on one of the analyses.

6 Interim analysis:

There will be no interim analysis.

7 Timing of analyses

Once all data for the last included patient have been obtained and the study team have performed final data cleaning, a final blind review and dummy-analysis will be carried out (with allocations completely made up) and any resulting updates or amendments to the statistical analysis plan (SAP) will be recorded. The final SAP will be signed off following this blind review/dummy analysis. This blinded review/dummy analysis will be invaluable as this is an atypical complex study design (stratified three-way crossover) with a novel analysis approach, so does not follow the template for analysis of more traditional trial designs. Any potential omissions/clarifications will be important to pick up prior to unblinding. As this is a three-way crossover trial, before unblinding, there will be no way of identifying the drugs patients receive as data will only be available by treatment period, and each treatment period consists of a mix of patients on all 3 drugs.

Unblinding of the data and final analysis on the unblinded dataset will not be performed until the dataset is documented as meeting the cleaning and approval requirements in accordance with the Trimaster Data Management plan, the database is locked, and the final statistical analysis plan (SAP) is fixed and approved. There will be two steps to unblinding: a coded unblinding (where the drugs are coded as A, B, and C but not revealed which is which), followed by full unblinding once the report of the coded blinded analysis has been written (See section 17).

8 Statistical principles

The mean, standard deviation, and any other summary statistics, will be reported to one decimal place greater than the original data. 95% confidence intervals will be reported for main effects. Main emphasis will be on 95% CIs but statistical significance will be assessed using $p<0.05$. Actual p values will be reported except for p values less than 3 decimal places which will be reported as "<0.001". P values >0.1 will be reported to one decimal place, p values between 0.001 and 0.1 will be reported to two significant digits.

Assumptions of statistical approaches/models will be assessed (e.g. normal distribution of HbA1c), and transformations used if necessary (although we do not anticipate the data violating model assumptions).

9 Important protocol deviations:

Participants will not be excluded from analysis based on protocol deviations alone. For protocol deviations leading to gaps in taking medication or delays in return visits, inclusion will be judged based on whether patients meet the necessary adherence criteria for inclusion based on proportion of tablets taken (see section 10). Failure to fast at the study visits will not affect HbA1c, tolerability and patient preference so this will not lead to exclusions for these main analyses. Cases where sample problems led to an HbA1c result not being available (e.g. difficulty bleeding the patient, remote visits due to coronavirus) will mean that study period will be excluded from primary analysis. Number of cases where this occurs will be recorded

10 Adherence to drugs

Medication adherence for each study period will be assessed based on the number of tablets returned. Participants are provided with 18 weeks worth of tablets (i.e. 126 tablets total) for each treatment period. At the end of each treatment period the participants return any unused medication and the number of tablets remaining is recorded. Adherence will be calculated as:

$$\frac{\text{Number of tablets actually taken}}{\text{Number of tablets expected to be taken}}$$

where the number of tablets actually taken within the treatment period will be calculated by 126 – number of tablets remaining at study visit, and the number of tablets expected to be taken (assuming 100% adherence in a treatment period) will be calculated by the total number of days in that treatment period (i.e. between study visits). In the MASTERMIND pilot study using MEMS-caps for monitoring medication adherence, pill count was shown to be as reliable as other patient reported measures. Patients with adherence <80% for diabetes therapies has been shown to affect glycaemic response[6]. Therefore, any participants with <80% calculated adherence for a given treatment period will be excluded from the treatment period concerned in primary analysis (but will be included in tipping point analysis and analysis of tolerability/patient preference). In cases, where patients did not bring their unused

tablets to the study visit, adherence for that study period will be based on self-reported compliance instead. Patients are asked four questions (if they ever forget to take their medicine, if they are careless about taking their medicine, if they stop taking their medicine if they feel unwell, if they stop taking their medicine if they feel better). We will consider patients to be non-adherent if they answer yes to at least three out of the four questions. The number of cases defined as non-adherent by each definition will be recorded.

11 Inclusion/Exclusion criteria of the overall Trial population (see section 13 for definitions of analysis-specific populations):
See main study protocol.

Patient disposition with numbers screened, randomised, and followed up with withdrawals will be presented in the form of a CONSORT diagram. The number and percentage of patients who completed the trial, discontinued study medication and who withdrew from follow-up will be presented for the whole randomised cohort in three tables: 1) overall, 2) broken down by each visit, and 3) for each treatment group. We will also present final sample sizes contributing to each of the two primary per-protocol analyses, and those of the full cohort for the tipping-point analyses and secondary analyses of tolerability and patient preference.

12 Withdrawals
The numbers of participants withdrawing at each stage, the drug(s) taken until withdrawal, and the reasons for withdrawal provided will be described. Differences in baseline characteristics (mean (SD) age, BMI, eGFR, HbA1c, and proportion male/female) between those withdrawing and those not will be reported.

13 Analysis populations:
There will be 3 separate datasets:

A) Main dataset

This dataset comprises baseline data and data from all treatment visits including on-treatment HbA1cs, and patient preference ranking but no additional information on side effects or adverse events. This will be used for the following:

1) Full cohort for CONSORT diagram – includes all screened and randomised participants regardless of whether they withdraw or are subsequently excluded.
2) Full cohort of all randomised participants.
3) Primary analysis cohorts – per-protocol analysis (**unblinded**):
   a) Hypothesis 1: TZD v DPP4i - Full cohort with the following participants excluded: withdrawal or early stopping before completing at least 12 weeks on TZD therapy; withdrawal or early stopping before completing at least 12 weeks on DPP4i therapy; adherence <80% on TZDi therapy; adherence <80% on DPP4i therapy. SGLT2 treatment period is not needed for this analysis. Baseline BMI and HbA1c at end of each treatment period must be recorded.
   b) Hypothesis 2: SGLT2i v DPP4i - Full cohort with the following participants excluded: withdrawal or early stopping before completing at least 12 weeks on SGLT2i therapy; withdrawal or early stopping before completing at least 12 weeks on DPP4i therapy; adherence <80% on SGLT2i therapy; adherence <80% on DPP4i therapy. TZD treatment period is not needed for

this analysis. Baseline eGFR and HbA1c at end of each treatment period must be recorded.

4) Primary analysis cohorts – per protocol analysis (**coded unblinding**)
   a) Hypothesis 1: same as hypothesis 1 for the unblinded analysis but there will be three separate cohorts as the analysis will be repeated 3 times (A v B, B v C and A v C) as it will not be known which of drugs A, B, or C are TZD and DPP4i.
   b) Hypothesis 2: same as hypothesis 2 for the unblinded analysis but there will be three separate cohorts as the analysis will be repeated 3 times (A v B, B v C and A v C) as it will not be known which of drugs A, B, or C are SGLT2i and DPP4i.
5) Tolerability analysis and tipping point analysis cohort: Same as full cohort of all randomised participants
6) Patient preference cohort: All participants who tried all 3 therapies and provided a ranking of their preference of the three therapies

B) Data from patient preference questionnaires including side effects.

7) Patient questionnaires data: The patient questionnaire data table will not be made available for primary analysis. Following primary analysis, this dataset will be merged with the full randomised cohort (dataset 2 described in part A above) to allow main analysis of side effects.

C) Adverse events data.

8) Adverse events data: The adverse events data table will not be made available for primary analysis. Following, primary analysis, it will be merged with the full cohort of all randomised participants (dataset 2 described in part A above) and used for reporting of safety.

14 Analysis

14.1 Primary Endpoint Analysis Definition

For the two main hypotheses, the aim is to determine whether individuals in a particular strata have a better glycaemic response to one drug compared with another. The primary outcome measure will be HbA1c at 4 months on each of the drugs, as a reasonable way to capture glycaemic control over this period.

For each treatment period, patients are provided with 18 weeks worth of the designated drug (126 tablets) to ensure adequate supply and allow flexibility for patients in booking their return visit. The primary outcome measure is HbA1c at 4 months, however when a participant withdraws from a treatment arm/discontinues treatment before 4 months, an HbA1c will still be taken and used in primary analysis if the subject has been taking the medication for at least 12 weeks (84 days). If the participant withdraws before taking the medication for 12 weeks, then the HbA1c for that period will be excluded from the primary analysis.

The HbA1c primary outcome measure will therefore be the achieved HbA1c after at least 12 weeks (≥84 days) on therapy. Where participants have <80% adherence (see section 10) on a drug, the HbA1c for that study period will be excluded from analysis.

A sensitivity analysis will be performed to determine the effect of an upper limit as some patients were provided with continuation bottles due to the covid pandemic (see section 15.2.1(iii)). A sensitivity analysis will also be performed to determine the effect of including those with an HbA1c taken at 12-15 weeks, rather than completing the full

4 month study period (defined as at least 15 weeks (≥105 days) to allow for flexibility in arranging the patient's study visit) (section 15.2.1(ii)).

This study does not aim to assess the efficacy of the medications. Instead, in line with our main hypotheses, we are interested in *if* a patient is able to tolerate the therapy, are they more likely to respond to a certain therapy given the strata they are in (see section 15.2 for full details including justification of analysis and additional sensitivity analyses). We will capture those who do not tolerate the therapy in our secondary endpoints (section 14.2).

14.2 Secondary Endpoint Definitions

1. Tolerability. Tolerability is when a patient takes a full course of medication. Patients have the option to discontinue their medication early and move onto the next drug (or end the study if they are on their third drug) if they are not tolerating the therapy well. We will consider patients as not tolerating the drug if they do not complete at least 12 weeks on that drug (i.e. 84 days) on therapy). A binary variable will be coded up for each treatment period with a code of 1 for if the patient completes at least 12 weeks (84 days) on the therapy and 0 if the patient does not (defined based on date of last study drug taken - date of prior study visit). Adherence is not considered in this endpoint definition.
2. Side effects. Patients will report at the end of each treatment period whether they have experienced any of the common side effects according to a list provided. There are 16 common side effects asked about in the patient questionnaire at the end of each study. A side effect will be considered to be associated with that study period if the patient answers "yes" to experiencing the side effect and it was not reported either at baseline or in any of the previous study periods. For weight and hypoglycaemia, we will provide quantification by using the weight (in kg) at the end of each treatment period and the self-reported frequency of hypoglycaemia (this is recorded as an ordinal variable: daily, weekly, monthly, occasionally). We will consider side effects reported in the patient questionnaire regardless of time on drug, as side effects may lead to discontinuation.
3. Patient preference. At the end of the study, each patient ranks the drugs they have taken in order of preference. At the end of each treatment period, HbA1c and weight are measured, and patients fill in a questionnaire reporting any side effects, frequency of hypoglycaemia episodes, and free text on their experience of taking the therapy. Patient preference will be based on their ranking of treatment preference <u>after</u> being fed back these data for each of their treatment periods (i.e. the second patient preference measure, although the first patient preference (before being informed of HbA1cs) will be used if the second preference is missing). In cases where there are ties:
    - if all 3 are considered equal, code all as rank 2
    - if prefer 1 and 2 equally, but 3rd least favourite, code as 1.5, 1.5 and 3
    - if prefer 1 but no preference for other 2, code as 1, 2.5 and 2.5

    We will only analyse data on preference in individuals who have taken all 3 drugs, but will not impose a minimum time on treatment for inclusion, as preference may be related to tolerability.

15 Statistical Analysis

15.1 Assessment of carryover and period effects

Prior to undertaking the main analysis, we will determine whether there is any evidence of carryover or period effects. For carryover, as this is a 3 period, 3 treatment study, we will examine first-order carryover effects (i.e. carryover from the preceding period only). These will be assessed by fitting a mixed effects model with carryover modelled by adding a term into the mixed effects model as follows[7]:

HbA1c ~ drug + period +carryover, random ~1|patient

(In Stata: mixed hba1c i.drug i.period i.carryover || patient:)

Where:

- **HbA1c** is the outcome (see section 14.1),
- **period** is the study period (1, 2, or 3), added as a factor
- **drug** is the drug (either TZD, DPP4i or SGLT2i, or coded as A, B, or C for coded analysis), added as a factor
- **carryover** is the drug they were on in the previous period (TZD, DPP4i or SGLT2, or coded as A, B, or C). Any level can be assigned to the carry-over variate in the first period, provided the same level is always used. Adjustment for periods then "removes" this part of the carry-over term[7].
- **Patient** is the patient ID added in as a random effect.

As period, drug, and carryover are all essentially dummy variables, the coefficients will represent change from the reference. With period 1 as the reference, the coefficients for periods 2 and 3 will determine the period effects for 2 and 3 compared with period 1. With drug A as the reference, the coefficients for carryover effects for drugs B and C will determine the carryover effect compared with drug A.

Overall, period and carryover can also be captured by ANOVA to determine whether there is *any* overall evidence of difference in HbA1c by period and by previous drug independent of period (carryover). *(In Stata: anova hba1c period carryover drug patient)*

Any statistically significant carryover effect identified (as defined by $p<0.05$) will be reported but not adjusted for in subsequent analysis[8]. Period effects will be reported and adjusted for in subsequent analyses.

15.2 Main analysis:

15.2.1 Choice of main analysis:

This is primarily a study of the effectiveness of 2 types of stratification rather than a study of each drug's efficacy.

The primary objective is to test two main hypotheses:

a) Obese patients (BMI >30kg/m^2), compared to non-obese patients, will achieve a lower HbA1c when assigned pioglitazone rather than sitagliptin.

b) Patients with an eGFR 60-90 mls/min/1.73m^2 will achieve a lower HbA1c, compared to patients with an eGFR>90 mls/min/1.73m^2, when assigned a sitagliptin rather than canagliflozin.

The key primary analysis for each hypothesis will be to assess whether the difference in achieved HbA1c for the two drugs differs for the two strata/groups of patients (see Table 1).

Patient group	**Drug A**	**Drug B**	**Difference**
In strata (S)	$HbA1c_{SA}$	$HbA1c_{SB}$	$HbA1c_{SA}$ - $HbA1c_{SB}$
Not in strata (N)	$HbA1c_{NA}$	$HbA1c_{NB}$	$HbA1c_{NA}$ - $HbA1c_{NB}$

Table 1 – stratum specific results and contrasts in final analysis

For a participant to be included in the primary analysis for each hypothesis, they will therefore need to have been on each of the two therapies for the given hypothesis for at least 12 weeks (see section 14.1). For this reason, our primary analysis will be a per-protocol approach.

Participants who are unable to tolerate therapy may also be informative so it will be important to consider their outcomes as well. However, intention to treat analysis requires some form of imputation of missing values. This is more challenging in a crossover setting because parallel group approaches such as imputing with the baseline are not valid as the pre-treatment baseline is only an appropriate baseline for the first period. Therefore, we propose two further analyses to explore the extent to which the missing HbA1cs could affect the final results: 1) an analysis of tolerability (see section 15.3) and 2) a tipping point analysis (see section 15.2.3(iv)).

15.2.2 Primary analysis: completers analysis

To begin with, the treatment contrasts will be calculated (HbA1c for drug A - HbA1c for drug B) and also the stratum specific contrasts (e.g. from Table 1 ($HbA1c_{SA}$ - $HbA1c_{SB}$) and ($HbA1c_{NA}$ - $HbA1c_{NB}$)). For hypothesis 1, strata will be obese ($BMI>30kg/m^2$) compared with non-obese ($BMI\leq30kg/m^2$) and drug A and drug B will be TZD and DPP4i. For hypothesis 2, strata will be eGFR 60-90 mls/min/1.73m^2 compared with eGFR>90, and drug A and drug B will be SGLT2i and DPP4i. For coded blinded analysis, 3 comparisons for each strata will need to be performed as it will not be known which of the 3 drug codings relate to SGLT2i, DPP4i, or TZD.

The null hypothesis is that the difference in achieved HbA1c for the two drugs is the same for the two strata i.e. ($HbA1c_{SA}$ - $HbA1c_{SB}$)= ($HbA1c_{NA}$ - $HbA1c_{NB}$). Standard errors for each treatment contrast and for each of the interactions of drug and strata will be calculated and used to construct 95% confidence intervals around the mean estimate.

As a first step, for each of the two main hypotheses, the difference in treatment contrasts between the two strata ($HbA1c_{SA}$ - $HbA1c_{SB}$ and $HbA1c_{NA}$ - $HbA1c_{NB}$) will be compared using a standard t-test.

The main analysis for each of the two hypotheses will then be undertaken using the following random effects models:

Hypothesis 1:

HbA1c ~ drug + obesity + period + (drug:obesity), random ~1|patient

which will compare the strata (obese v non-obese) on the two drugs of interest (pioglitazone v sitagliptin). The key contrast of interest is the drug:obesity interaction

as this is the difference in treatment contrasts between the obese and non-obese strata.

This model will be carried out in Stata using the command

mixed hba1c drug obese i.period drug##obese || id:

where drug is a binary variable (TZD v DPP4i), obese is a binary variable (1 if obese, 0 if not), period is a factor with 3 levels, drug##obese is the drug*obesity interaction. The coefficient, 95% confidence intervals and p value will be presented for the drug*obesity interaction.

Hypothesis 2:

HbA1c ~ drug + EGFR group + period + (drug:EGFR group), random ~1|patient

which will compare the strata (EGFR 60-90 v EGFR>90) on the two drugs of interest (canagliflozin v sitagliptin). The key contrast of interest is the drug:EGFR group interaction.

In line with hypothesis 1, this model will be carried out in Stata using the command

mixed hba1c drug egfrgp i.period drug##egfrgp || id:

where drug is a binary variable (DPP4i v SGLT2i), egfrgp is a binary variable (1 if eGFR60-90, 0 if eGFR>90), period is a factor with 3 levels, drug##egfrgp is the drug*egfrgp interaction. The coefficient, 95% confidence intervals and p value will be presented for the drug*egfrgp interaction.

The two hypotheses will be tested separately.

A table with all beta-coefficients, 95% confidence intervals, z values and p values will be presented.

15.2.3 Pre-specified sensitivity analyses:

i) Additional analysis accounting for change to inclusion criteria

On 1st August 2017 the trial inclusion criteria changed (protocol v5). Prior to this date, patients were only eligible if they were treated with both metformin and sulphonylureas. This was broadened to allow inclusion of patients who were treated with metformin alone. To determine whether this change in inclusion criteria impacts on the primary analysis, we will also carry out an additional analysis, for each hypothesis, adjusting for study “epoch” which will be coded as 1 before the inclusion criteria change and 0 for after the inclusion criteria change.

HbA1c ~ drug + strata + period + (drug:strata) + epoch, random ~1|patient

This will result in two models being be carried out in Stata similar to those reported in 15.2.2, where strata is either obese (for model 1) or egfrgp (for model 2), with the additional term for epoch added in.

The beta coefficient, 95% CI, and p value for the epoch term will be reported, along with any change in the drug:strata interaction coefficient compared with that seen in the main primary analysis.

ii) Analysis of only patients completing the full treatment period
We will repeat the main analysis for each hypothesis but only including participants who complete at least 15 weeks (105 days) on both drugs for that hypothesis.

iii) Analysis examining impact of COVID-19 protocol amendment
On 25th March 2020, the trial protocol was amended to extend visit windows and provide an additional "continuation" bottle of therapy to allow more flexibility for patients in terms of their return visit and to ensure that they had adequate supply to continue on therapy (protocol version 8). This means that the length of time participants receive the study drug could be longer. We will investigate whether receiving the study drug for >18 weeks (>126 days) impacts on the main study hypothesis:

HbA1c ~ drug + strata + period + (drug:strata) + gt18wk, random ~1|patient

where gt18wk is a binary category coded as 1 if the treatment period is >18 weeks and 0 if the treatment period <=18 weeks. The beta coefficient, 95% CI, and p value from the gt18wk term will be reported, along with any change in the drug:strata interaction coefficient compared with that seen in the primary analysis.

iv) Tipping point analysis
A tipping point analysis will be used to explore what change in treatment contrast would be required as a result of the missing data to significantly change the outcome. A tipping point analysis can be considered as a kind of meta-analysis of two studies including the observed results from the completers study and the hypothetical results if data were available on those for whom we have missing values.

In the main completers analysis, the treatment contrast is defined as:

$$\tau = (HbA1c_{SA} - HbA1c_{SB}) - (HbA1c_{NA} - HbA1c_{NB})$$

The treatment contrast for a study involving values from all participants is not known but we can assign a hypothetical value, Δ, to represent the difference between the treatment contrast from the completers analysis and what the treatment contrast would be if we had values for everyone:

$$\tau_{comb} = \tau + \Delta f$$

where τ is the treatment contrast from the completers analysis and f is the fraction of the cohort with missing data (where full dataset is all patients with baseline data from visit 1).

T-statistics for a range of delta values will be calculated using:

$$t = \frac{\tau + \Delta f}{SE}$$

where τ is the treatment contrast for completers analysis, f is the fraction with missing data and SE is the standard error for τ calculated as the observed standard deviation of the treatment contrast from the completers analysis divided by the square root of total n for the whole cohort including missing data.

A plot of calculated t-statistic values against delta values will be produced.

The tipping point will be designated according to when it will change the outcome at the 5% significance level and calculated by:

$$\Delta = \frac{(1.96 \cdot SE) - \tau}{f}$$

Note that τ could be positive or negative depending on which way the difference is calculated so calculate Δ for both.

15.3 Analyses of the secondary outcomes:

15.3.1 Tolerability

15.3.1.1 – Tolerability overall
As an initial simple analysis, we will report the proportion who tolerate each therapy (i.e. complete at least 12 weeks on the therapy, see 14.2.3).

It is possible that one of the drugs may be more likely to lead to drop out. To allow for this, we will first use the Mantel-Haenszel approach, where for each period we will produce contingency tables of drug against tolerability. These data can be analysed as a meta-analysis using the Mantel-Haenszel procedure to calculate the overall common odds-ratios (and 95% confidence intervals) for each drug compared with the other two across all three periods.

We will then perform further analysis using mixed effects logistic regression, with drug and period as fixed effects and patient as a random effect:

Tolerability ~ drug + period, random=~1|patient

In Stata, mixed effects logistic regression will be carried out using the xtmelogit command:

xtmelogit tolerability i.drug i.period || patient:, or

- where tolerability is a binary variable coded as 1 for those who tolerate therapy, 0 as those who do not (see section 14.2).
- drug is a factor with three levels (TZD, SGLT2i, DPP4i)
- period is a factor with three levels (1, 2, 3)
- patient is the individual id set as a random effect

A table presenting the beta coefficients (converted to odds ratios), corresponding 95% confidence intervals, z values and p values will be presented. The key coefficients will be the odds ratio (and 95% confidence intervals) for tolerating therapy for each drug compared with the reference drug.

15.3.1.2 – Tolerability by strata
We will also examine whether tolerability differs by each of the strata for the different drugs. As a first step, we will compare the proportion tolerating each therapy in each strata (obese/non-obese or eGFR60-90/eGFR>90).

We will then examine this further using mixed effects logistic regression models:

Tolerability ~ drug + strata + period + (drug:strata), random=~1|patient

In line with the primary analysis, this will consist of two models: 1) investigating the obese/non-obese strata for the DPP4i and TZD treatment arms, 2) investigating the eGFR 60-90 v eGFR>90 strata for the DPP4i and SGLT2i treatment arms. The null hypothesis for each analysis will be that the difference in tolerability between the two drugs is the same.

Each hypothesis will be tested separately. In Stata this will result in two models:

1) Hypothesis 1:
   xtmelogit tolerability drug obese i.period drug##obese || patient:, or
   where drug will be a binary variable (TZD or DPP4i)

2) Hypothesis 2:
   xtmelogit tolerability drug egfrgp i.period drug##egfrgp || patient; or
   where drug will be a binary variable (DPP4i or SGLT2i)

As before, the key coefficient of interest will be the drug##obese or drug##egfrgp interaction. For each hypothesis, a table of all beta coefficients (converted to odds ratios), corresponding 95% confidence intervals, z values and p values will be presented.

15.3.1.3 – Tolerability – sensitivity analysis

As a sensitivity analysis, we will also undertake analysis defining tolerability based on completing at least 15 weeks on therapy (i.e. not stopping the therapy early), as discontinuing after 12 weeks and not completing the full study period may still indicate not tolerating the therapy. This will involve performing the same analyses as *15.3.1.2* but instead defining the outcome of tolerability as a binary variable coded as 1 if completed at least 15 weeks on therapy (>=105 days), 0 if not.

We will carry out a largely descriptive exploration of the reasons for intolerance, including comparison of frequency of reported side effects in those tolerating/not tolerating therapy.

15.3.2 Side effects

We will examine the distribution of side effects reported across each of the 3 drugs. Given the total numbers reporting each individual side effect will likely be small, this will largely be descriptive, examining the side effect profile observed with each drug. There are 16 common side effects that patients are asked to report if they have experienced as part of the questionnaire at the end of each study period.

Side effects will be defined as <u>new</u> side effects only (i.e. not reported at the baseline visit or in previous treatment periods). Summary statistics will be presented showing the number of participants reporting each side effect for each drug (and percentage of total receiving that drug). This will then be subdivided by questions in the patient questionnaire which asks for each side effect

- Did you see a doctor?
- Did you receive any treatment?
- Did this interfere with your everyday life?

Similar mixed effects models to those used in primary analysis will be employed to assess the secondary endpoints of weight at 4m and frequency of hypoglycaemia in the last month on each drug (with ordinal mixed effects models used for frequency of hypoglycaemia). However, it will be important to note that the study was not powered for this analysis, nor was the study designed to consider potential carryover for either of these outcomes. Therefore, power for the final sample size and carryover and period effects will be reported for both of these outcomes.

15.3.3 Patient preference

For this analysis, we will only analyse the dataset where the participants have tried all 3 drugs. The mean rank for each drug will be calculated. The null hypothesis is that there is not a preferred drug and therefore the expected value of the rank for a given drug will be 2. Statistical significance of departure from the null hypothesis will be determined using Z-statistics constructed from the mean rank R and its variance over n patients (2/(3n)), where:

$$Z = \frac{R-2}{\sqrt{2/(3n)}}$$

Further investigation of predictors of patient preference will be largely exploratory.

15.3.4 Overall HbA1c by drug class regardless of strata

To determine whether there is a difference between drug classes in terms of the overall achieved HbA1c after 4 months on each of the drugs, we will also fit the following model:

HbA1c at 4 months ~ drug + period, random=~1|patient

Drug will be a factor and coded as a dummy variable as the comparison will be across all 3, rather than 2, drug classes. Based on this model, mean and 95% CI for HbA1c on each therapy, adjusted for period, will be reported.

16 Missing data:

To be eligible for at least one of the two primary analyses, the participant will need to have HbA1c results for at least two of the drugs, one of which being a DPP4-inhibitor. For any period where the patient discontinues the drug within 12 weeks, the HbA1c for that period will be excluded. There will be no imputation of missing HbA1c results (see section 15.2).

Any patients with missing baseline eGFR or BMI will be excluded from analyses of the respective strata.

17 Blinded Analysis after data lock

The first analysis following data lock will be performed blinded to treatment allocation. The study database team will provide a coded unblinding where treatment allocation will be given to the study statistician but coded as A, B, or C, but not revealing which code relates to which drug. There will therefore be six separate analyses undertaken whilst blinded: three separate cohorts (A v B, B v C, and A v C) for hypothesis 1 (testing drug:obesity interaction) and, similarly, three cohorts for hypothesis 2 (drug eGFR group interaction). Note – only one of each of the 3 analyses will be relevant on unblinding and the other comparisons will be ignored. Blinded analysis using this coded dataset will also be performed for patient preference, side effects, and tolerability outcomes. Analysis will be carried out in parallel by the study statistician and an independent statistician from the Exeter Clinical Trials Unit. Discrepancies in analysis and final results will resolved by discussion and details recorded. Final drug coding will be unblinded after the report of the first blinded analysis has been produced and submitted to the Trial Steering Group.

18 Statistical software

Stata v16.1 will be validated and used as the primary statistical software.

19 Amendment history:

Amendment No.	Protocol version no.	Date issued	Author(s) of changes	Details of changes made
1	V6	15/05/2018	Catherine Angwin/Beverley Shields	Following discussion with Steering Committee and Data Monitoring Committee, amendment to sample size due to over-cautious calculations (alpha changed to 0.05 from 0.01).
2	SAP v1	15/01/2019	Beverley Shields	Taken out of study protocol and put into a separate statistical analysis plan. Additional info added on dealing with withdrawals/missing data/protocol violations and the plans for secondary analysis of patient preference expanded.
3	V7	22/02/2019	Catherine Angwin/Beverley Shields	To allow for early drop outs (estimated at 19%), sample size increased to 520.
4	SAP v2	16/08/2019	Beverley Shields	Revisions to Statistical Analysis Plan following discussion with Steering Committee statistician. Concerns over using baseline in those who stop therapy early. Changed from modified intention to treat analysis to per protocol type analysis and added in separate tolerability outcome to explore those stopping early. Updated to include modified sample size.
5	SAP V3	Feb 2020	Beverley Shields	Provided more detail throughout SAP. Following advice from MASTERMIND consortium and discussed with Steering Committee, changed outcome from using HbA1c if at least 8 weeks on therapy to using HbA1c if at least 12 weeks on therapy (biological rationale). On advice of Steering Committee statistician, added simplified patient preference analysis and added in more basic analysis to check robustness of more complex modelling.
6	SAP V4	July 2020	Beverley Shields	Adding in amendments relating to COVID changes.
7	SAP V5	Oct 2020	Beverley Shields	Amendments following review from head of CTU
8	SAP v6	Jan 2021	Beverley Shields	Final analysis added in following discussion with Stephen Senn and Rury Holman (tipping point sensitivity analysis, and Mantel-

				Haenszel analysis for tolerability). Gender hypothesis removed as unlikely to be powered (male/female split not even; ~75% male) and not a main hypothesis.
9	SAP v7	09/02/2021	Beverley Shields	Further clarification of sections that were not clear, justification for blind review/dummy run added in.
10	SAP v8	17/02/2021	Beverley Shields	Following dummy analysis on **simulated** data in Stata (not on actual data). Further clarification of sections, models rewritten in Stata format (as the intended software package for analysis), amendment to model to assess carryover effects (as initial proposed model not sufficient for 3x3 crossover design).
11	SAP v9	11/03/21	Beverley Shields	Following blind review and dummy analysis on real data (but no allocations provided), minor changes where definitions required further clarification (e.g. tolerability does not include adherence in the definition, patient preference refers to the second set of patient preferences unless missing (in which case the first will be used)

20 References

1. Dennis JM, Henley WE, Weedon MN, et al. Sex and BMI Alter the Benefits and Risks of Sulfonylureas and Thiazolidinediones in Type 2 Diabetes: A Framework for Evaluating Stratification Using Routine Clinical and Individual Trial Data. *Diabetes Care* 2018;41(9):1844-53. doi: 10.2337/dc18-0344 [published Online First: 2018/08/04]
2. Dennis JM, Shields BM, Hill AV, et al. Precision Medicine in Type 2 Diabetes: Clinical Markers of Insulin Resistance Are Associated With Altered Short- and Long-term Glycemic Response to DPP-4 Inhibitor Therapy. *Diabetes Care* 2018;41(4):705-12. doi: 10.2337/dc17-1827 [published Online First: 2018/02/02]
3. Tahara Y, Shima K. Kinetics of HbA1c, glycated albumin, and fructosamine and analysis of their weight functions against preceding plasma glucose level. *Diabetes Care* 1995;18(4):440-7. doi: 10.2337/diacare.18.4.440 [published Online First: 1995/04/01]
4. Tahara Y, Shima K. The response of GHb to stepwise plasma glucose change over time in diabetic patients. *Diabetes Care* 1993;16(9):1313-4. doi: 10.2337/diacare.16.9.1313 [published Online First: 1993/09/01]
5. Lund SS, Tarnow L, Stehouwer CD, et al. Targeting hyperglycaemia with either metformin or repaglinide in non-obese patients with type 2 diabetes: results from a randomized crossover trial. *Diabetes Obes Metab* 2007;9(3):394-407. doi: 10.1111/j.1463-1326.2007.00713.x [published Online First: 2007/03/30]

6. 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* 2016;39(2):258-63. doi: 10.2337/dc15-1194 [published Online First: 2015/12/19]
7. Jones B, Kenward MG. Design and analysis of cross-over trials. 2nd ed. ed. Boca Raton, Fla. ; London: Chapman & Hall/CRC 2003.
8. Senn S. Cross-over trials in clinical research. Chichester: Wiley 2002.