The **G**lycemia **R**eduction **A**pproaches in **D**iabetes: A Comparative **E**ffectiveness **Study (GRADE Study)**

NCT01794143

Statistical Analysis Plans for Manuscript Entitled "Glycemia Reduction Approaches in Type 2 Diabetes: A Comparative Effectiveness (GRADE) Study Microvascular and Cardiovascular Outcomes" SAP

Initial Statistical Analysis Plan

May 12, 2021

Final Statistical Analysis Plan

April 8, 2022

Summary of Changes to the Statistical Analysis Plan

Sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)

GRADE Study Coordinating Center
Biostatistics Center
George Washington University
6110 Executive Boulevard
Rockville, Maryland 20852

Statistical Analysis Plans for Manuscript Entitled "Glycemia Reduction Approaches in Type 2 Diabetes: A Comparative Effectiveness (GRADE) Study Microvascular and Cardiovascular Outcomes"

Table of Contents

Treatment group differences in micro/macrovascular outcomes among four initial treatments added to metformin in early type 2 diabetes (OP2)

Table of Contents

1 GENERAL INFORMATION

GRADE paper number	OP 2
Analysis Category	End of Study
Writing Group Chairs	David Nathan, John Lachin
Writing Group Members	John Lachin, David Nathan, Henry Burch, Andrea Cherrington, Stephen Fortmann, Jennifer Green, Sue Kirkman, Lawrence Phillips, Rodica Pop-Busui, Michael Steffes, Margaret Tiktin
Target Journal	NEJM
Lead Statisticians	Insert Name

2 APPROVALS

	sign off on	**Signature**
David M. Nathan GRADE Co-PI and Writing Committee Co-Chair	Aims, outcomes, mock tables and figures	
John M. Lachin GRADE Co-PI and Writing Committee Co-Chair	Aims, outcomes, mock tables and figures, statistical methods	
Naji Younes Supervisory Statistician	Aims, outcomes, mock tables and figures, statistical methods	
Mark Tripputi Analytic Statistician	Aims, outcomes, mock tables and figures, statistical methods	

3 REVISION HISTORY

Version No.	Implemented by	Date	Reason
1	Mark Tripputi	05/02/2021	Initial version
2	Mark Tripputi	05/06/2021	Revisions from John's Review
3	Mark Tripputi	05/08/2021	Revisions from John's Review

4 ABBREVIATIONS AND ACRONYMS

Abbreviation	**Meaning**

GLP-1	Glucagon Like Peptide 1
DPP-4	Dipeptidyl peptidase 4
SGLT-2	sodium-glucose cotransporter 2
MACE	Major adverse cardiovascular events
CHF	Congestive Heart Failure
DSPN	Distal symmetric polyneuropathy
eGFR	estimated glomerular filtration rate
UACR	Urine Albumin-to-Creatinine Ratio
RMST	Restricted Mean Survival Time

5 STUDY OBJECTIVES

5.1 Background and justification

Type 2 diabetes (T2DM) affects more than 30 million persons in the United States, with an incidence of 1.5 million new cases per year, and more than 400 million persons world-wide. The major human and economic costs associated with T2DM are related primarily to the development of long-term diabetes-specific complications, including retinopathy, nephropathy, and neuropathy, and a 2-5 fold increased risk of non-specific cardiovascular disease (CVD). These long-term complications have been shown to be ameliorated in part by interventions that reduce chronic glycemia, as measured by glycated hemoglobin levels (HbA1c), and a target range of less than 7% (53 mmol/mol) has been established by consensus for most patients with T2DM. The estimated annual cost of diabetes in the US in 2017 was approximately $327 billion dollars per year with an increasing fraction attributed to the cost of glucose-lowering medications.

Virtually all recommendations for the management of type 2 diabetes have included metformin as the first medication to be used. Unfortunately, choosing the second medication from the ever expanding list of glucose-lowering medications to add to metformin when monotherapy fails to achieve or maintain goal glycemia is problematic owing to the dearth of any long-term head-to-head comparator studies. The purpose of the Glycemia Reduction Approaches in Type 2 Diabetes: A Comparative Effectiveness (GRADE) Study was to examine the relative effectiveness of the four most commonly used glucose-lowering medications added to metformin to maintain goal glycemia. In this paper, we report the difference between the treatment groups in the incidence of micro- and macrovascular outcomes and their risk factors. The accompanying paper reports treatment group differences in the metabolic outcomes.

5.2 Scientific objectives

1. Compare cumulative incidence of outcomes by treatment group to assess whether one or more of the treatment approaches had increased (or decreased) benefit compared with the others

2. Compare the relative efficacy of treatment groups (hazard ratios) on the risk of the micro/macro vascular outcomes. This will include both pairwise comparisons, and comparisons of each group vs the other groups combined.

3. Compare profiles (baseline, year1, year5) of micro/macrovascular risk factors by treatment groups

4. Assess whether treatment group effects on micro/macrovascular outcomes are mediated by levels of HbA1c.

5. Compare differences in treatment group effects on micro/macrovascular outcomes by levels of relevant subgroup variables

6 STATISTICAL METHODS AND DATASETS

6.1 Analysis Data Set Inclusion Criteria

The study will include all GRADE randomized participants.

6.2 Outcomes to be Assessed

There are six micro/macrovascular outcomes of primary interest in this paper:

1. Major Adverse Cardiovascular Events (MACE): consists of non-fatal MI, stroke, or cv death
2. Heart Failure (HF)
3. Distal Symmetric PolyNeuropathy (DSPN)
4. eGFR < 60 $mL \cdot min^{-1} \cdot (1.73m^2)^{-1}$
5. Confirmed Microalbuminuria (UACR ≥ 30 $mg \cdot g^{-1}$ on 2 successive evaluations)
6. Macroalbuminuria (UACR ≥ 300 $mg \cdot g^{-1}$)

Definition of Major Adverse Cardiac Events(MACE)

- Defined as occurrence of one or more of the 3 MACE components: non-fatal MI, non-fatal stroke, cardiovascular death including fatal stroke or fatal MI
- Requires adjudication (Cardiovascular Event (MACE and Non-MACE) Adjudication Form)
- Do we want a Wei-Lachin analysis in addition to a composite?

Definition of Heart Failure(HF)

- Defined as diagnosis of HF requiring hospitalization with adjudication (Cardiovascular Event (MACE and Non-MACE) Adjudication Form)

Definition of DSPN

- DSPN will be assessed using the MNSI questionaire and neuropathy exam components (1)
- MNSI questionaire (collected annually)
  - 15 items with each question scored 0 if symptom absent, 1 if present
- Neuropathy exam (annual form) consists of 5 components assessing both left and right feet
  - Appearance (Normal/Abnormal)
  - Ulceration (Y/N)
  - Ankle reflexes (Present/Absent)
  - Vibration perception using tuning fork on top of great toes (absent, reduced, present)
  - Detection of 10g monofilament on top of great toes, with 10 trials on each toe (absent, reduced 1-7, present ≥ 8)
- DSPN assessments are conducted at baseline and annually during follow-up

The prevalence and incidence of DSPN will be defined on the basis of these assessments as follows:

- Incident or prevalent DSPN will be defined as an MNSI symptom score of ≥ 7.5 AND/OR MNSI exam score (for appearance, ulcer, reflex, vibration – excluding the monofilament) of ≥ 2.5, occurring among any participant who did not satisfy either of those criteria at baseline.

In addition, a report of any ulcerations or amputations of any part of foot or lower extremity are captured as SAEs.

Definition of eGFR < 60 outcome

- (eGFR < 60 $mL \cdot min^{-1} \cdot (1.73m^2)^{-1}$ and baseline value ≥ 60) OR ESRD death OR dialysis/transplant
- The eGFR measurement is based on the serum creatinine that is collected annually
- Dialysis OR Transplant reported as single outcome (quarterly and annual forms)
- ESRD death is adjudicated

Definition of Confirmed Microalbuminuria

- (UACR ≥ 30 $mg \cdot g^{-1}$ AND baseline value < 30) OR ESRD death OR dialysis/transplant
- UACR (collected every 6 months) threshold confirmed at two consecutive visits

Definition of Macroalbuminuria

- Protocol defined as occurrence of UACR ≥ 300 $mg \cdot g^{-1}$ and baseline value < 300 OR ESRD death OR dialysis/transplant

6.3 Assessment of Study Power

The GRADE protocol contains power calculations for MACE and confirmed microalbuminuria:

- With a projected incidence rate of 0.04/yr for *confirmed microalbuminuria* (indicates rate in group with highest incidence), GRADE would have 88% power to detect a 33% difference in risk between any of the 6 pairwise treatment group comparisons
- With a projected incidence rate of 0.01/yr for *MACE* (indicates rate in group with highest incidence), GRADE would have 80% power to detect a 50% difference in risk between any of the 6 pairwise treatment group comparisons. Smaller differences can be detected if each group is compared with the other 3 combined (80% power to detect a 42% difference in risk)

Event Counts

Event rates calculated using the **`grCore:/RC_3.0 events`** dataset (not the complete final dataset)

- Data for nephropathy events
  - table gives both unconfirmed (N) and confirmed counts (N confirmed) of participants with one or more events
  - confirmation was defined as occurrence of the events at two consecutive expected visits. For example, serum creatinine was collected at baseline and annually during followup, so "consecutive visits" was interpreted as two successive annual visits. Note that events would not confirm if there was a missing visit between two occurrences (e.g. event at 12mo, missing 24mo visit, event at 36mo would not count as confirmed).
  - eGFR is not a confirmed event.
- Data for MACE and CHF events
  - table gives the count of participants with one or more adjudicated MACE events (cv death OR non-fatal stroke OR non-fatal MI), the component events, and CHF events
- DSPN events still to be defined in the dataset

Event	N	N confirmed[1]	At Risk yrs	Annual Rate
Micro-albuminuria	1320	641	3.50, 4.05	0.075, 0.031

Macro-albuminuria	250	N/a	4.72	0.010
eGFR < 60	608	N/a	4.72	0.025
MACE	190	N/a	4.95	0.008
CV death	17	N/a	2.58	0.001
stroke	81	N/a	5.26	0.003
MI	107	N/a	4.94	0.004
DSPN	1722	N/a	5.22	0.065
CHF	103	N/a	3.11	0.0007

[1]Confirmation only applies to the microalbuminuria outcome N/a: not applicable

6.4 Statistical Analyses

All figures and tables referenced in this section are shown in sections below (Proposed Tables, Proposed Figures).

6.4.1 Scientific objective 1: Treatment effect on Cumulative Incidence

Objective Compare crude rates and cumulative incidence of outcomes by treatment group (Figure 1)

Statistical Analyses

Figure 1 description (see mockup below)

A 2x3-panel figure.

The 6 panels in the top row display the cumulative incidence for the MACE, CHF, DSPN, eGFR < 60, confirmed microalbuminuria, and macroalbuminuria outcomes (from left to right) over time. Each panel includes 4 lines, one for the cumulative incidence within each treatment group. The cumulative incidence by treatment group will be estimated using a Kaplan-Meier estimator. The total number at risk at each year will be provided below each panel. The time axis will represent the time since GRADE randomization. The maximum value for the time axis will be selected as the last time when the total number at risk is ≥ 200 for the outcomes. The unadjusted log-rank test will be used to assess differences between the treatment groups(2).

A simple mocked-up version of figure 1 is displayed in the section Proposed Figures

6.4.2 Scientific objective 2: Treatment effect on hazard ratios and RMST

Objective Compare risk of outcomes between treatment groups (Table 2)

Statistical Analyses

For this table, the following statistics will be calculated for the micro/macrovascular outcomes, both overall and stratified by treatment group:

- The number of events and percent of the GRADE cohort with the outcome.
- Crude rate per 100 person-years (SE). This will be calculated as 100*(observed number of events)/(total time at risk), where the total time at risk is the sum of the time since randomization to the event (or to the censoring time for those without an event) across participants.
- Pairwise hazard ratios (SE). A Cox proportional hazards model will be fit for the outcome with treatment group as a predictor. For the purposes of this Cox model, the event times and censoring

times will be calculated as time since randomization to the event or censoring, respectively. Hazard ratios and standard errors for each pairwise comparison of the treatment groups will be estimated from the Cox model. All Wald-type tests, standard errors and confidence intervals will be estimated using the robust (3) information sandwich estimator to ensure valid inferences even if the proportional hazards assumption does not apply. A joint test for differences in the hazards among any of the treatment groups will be conducted. If that joint test is significant, then pairwise log-rank tests will be conducted to test for all pairwise differences. There are a total of 6 possible pairwise comparisons among the 4 treatment groups, and therefore these tests will be adjusted for multiple comparisons using a closed testing procedure (see details in the Other statistical issues section of this document). If the joint test for differences among any of the treatment groups is significant, then the results from the pairwise testing will be visualized using the following graphic, where each corner of the box represents one of the four treatments (G = Glimepiride, L = Liraglutide, S = Sitagliptin, I=Insulin Glargine), and lines connect the treatments that differ significantly; dotted lines indicate $p \leq 0.05$, dashed lines indicate $p \leq 0.01$, and solid lines indicate $p \leq 0.001$.

- Hazard ratio compared to all other treatments combined (SE) (4). A Cox proportional hazards model will be fit for the outcome with treatment group as a predictor. For the purposes of this Cox model, the event times and censoring times will be calculated as time since randomization to the event or censoring respectively. For a given treatment group, the hazard ratio compared to all other treatments combined will be estimated as the average of the estimated hazard ratios comparing each of the other treatments to the given treatment group. Since there are 4 treatment groups, there would be a total of 4 tests comparing each treatment to all others combined, and therefore these tests will be adjusted for multiple comparisons using the previously described closed testing procedure (see details in the “Other statistical issues” section at the end of this document).
- Pairwise RMST ratios (SE). A log-linear model will be fit for the restricted mean survival time (RMST) up to time = 4 years using inverse probability of censoring weighting (IPCW) (Tian et al, 2014). RMST ratios and standard errors for each pairwise comparison of the treatment groups will be estimated from this model. The same testing procedure for the pairwise comparisons will be used as for testing pairwise hazard ratios above.
- RMST ratio for each treatment compared to all other treatments combined (SE). A log-linear model will be fit for the restricted mean survival time (RMST) up to time = 4 using inverse probability of censoring weighting (IPCW) (Tian et al, 2014). For treatment a, the RMST ratio compared to all other treatments combined will be estimated as the average of the estimated RMST ratios comparing each of the other treatments to treatment a. The same testing procedure will be used as for testing hazard ratios compared to all other treatments combined above.

6.4.3 Scientific objective 3: Risk factor profiles by treatment group

Objective Compare profiles (baseline, year1, year4) of micro/macrovascular risk factors by treatment groups (Table 3)

Risk factors of interest include UACR, eGFR, systolic blood pressure (SBP), diastolic blood pressure (DBP), HDL, LDL, triglycerides, and prevalence of eGFR < 60, micro/macro albuminuria, SBP > 140, use of blood pressure lowering medications, use of ACEi/ARBs, and use of statins.

Statistical Analyses

- GEE models will be used to estimate time averaged values of the risk factor with an unstructured covariance matrix for the multiple time point measurements. Inclusion of a time by treatment interaction terms will allow for testing of heterogeneity of time effects by treatment group. The models will be adjusted for baseline levels of the risk factor. The following estimates will be presented in the table:
  - mean values and standard errors of risk factors overall at baseline, 1yr and 4yr post-randomization
  - mean values and standard errors of risk factors in treatment groups at 1yr and 4yrs (assumed common baseline value from overall estimate)
  - the p-value from a test of interaction of time by treatment group will be used to assess if there is heterogeneity in the time trend of estimates by treatment group.

6.4.4 Scientific objective 4: Mediation Analyses

Objective Assess whether treatment group effects on micro/macrovascular outcomes are mediated by levels of HbA1c. (Table 4)

Statistical Analyses

Mediation analyses will be conducted to estimate the proportion of treatment effects on the micro/macro vascular outcomes that are explained by HbA1c as a mediator. This analysis will follow Baron and Kenny's mediation paradigm (5). First, an unadjusted model for the outcome by treatment group will be fit, and the treatment effect of each treatment vs. all others will be estimated from this model (θ_{0k} for treatment k). Then, a second model will be fit for the treatment effect on the outcome adjusted for the current value of HbA1c (i.e., HbA1c as a time-varying covariate), and the treatment effect of each treatment vs. all others will also be estimated from this model (θ_{1k} for treatment k). Finally, the percent mediation of the treatment effect by HbA1c for each treatment will be calculated as the relative change in the treatment effect in a model adjusted for HbA1c as a mediator relative to an unadjusted model (i.e., $([(\theta_{0k} - \theta_{1k}) \div \theta_{0k}] * 100\% = (1 - \theta_{1k}/\theta_{0k}) * 100\%)$.

Since the proportional hazards assumption is not preserved under marginalization (i.e., the proportional hazards assumption cannot hold for both the unadjusted model and the model adjusted for Hba1c as a mediator; (6)), standard errors will be estimated using a robust information sandwich estimator (3) to ensure valid inferences when the proportional hazards assumption does not apply.

6.4.5 Scientific objective 5: Subgroup Analyses

Assess if there are differences in treatment group effects on micro/macrovascular outcomes by levels of relevant subgroup variables (Table 5, Figure 2)

Statistical Analyses

Subgroup analyses of the treatment effects for the micro/macrovascular outcomes within subgroups based on the following baseline variables:

- age (< 45, 45-59, 60+)
- sex (male, female)
- race (Non-Hispanic white, non-Hispanic black, Hispanic white, and other)
- HbA1c (tertiles)
- diabetes duration in years (tertiles)

Hierarchical closed testing of subgroup by treatment group interaction will be used to identify subgroups within which some heterogeneity may exist, and within each such subgroup the treatment groups will be compared using a closed testing procedure (see details in the "Other statistical issues" section at the end of this document(7)). Results of will be presented in the overall group and in any subgroups with heterogeneity.

Table 5

- The number of participants in each treatment group within each subgroup.
- The number of events for each micro/macrovascular outcome in each treatment group within each subgroup.
- Crude rate per 100 person-years (with 95% confidence intervals) of each micro/macrovascular outcome in each treatment group within each subgroup. The crude rates will be calculated as 100*(observed number of events)/(total time at risk), where the total time at risk is the sum of the time since randomization to the event (or to the censoring time for those without an event) across participants.
- P-value from overall test of homogeneity of treatment effect across each baseline subgroup variable. For quantitative factors (i.e., diabetes duration, HbA1c), this p-value will be based on a test of homogeneity of treatment effect across the continuous quantitative variable (i.e., not based on the tertiles).
- Hierarchical closed testing of subgroup by group interaction will be used to identify subgroups within which some heterogeneity may exist, and within each such subgroups the treatment groups will be compared (7). Tests of pairwise treatment comparisons within a subgroup will be visualized in the same way as the tests of pairwise treatment comparisons within subgroups in Figure 2.

Figure 2 A 3x2-panel figure. There is a separate panel for each of the baseline subgroup variables. Tests of all pairwise treatment comparisons within each subgroup will be assessed. Since there are a total of 6 possible pairwise comparisons within each subgroup, these tests will be adjusted for multiple comparisons using a closed testing procedure (see details in the "Other statistical issues" section at the end of this document). The results from the pairwise testing within each subgroup will be visualized using the same graphic as described in Table 2. For subgroups that are determined not to have heterogeneous treatment effects, treatment effects will not be tested within subgroup, and so this graphic will be omitted.

Each panel will display the crude rates per 100 person-years (with 95% confidence intervals) of the primary outcome for each treatment group within each subgroup of the baseline variable. The crude rates will be calculated as 100*(observed number of events)/(total time at risk), where the total time at risk is the sum of the time since randomization to the event (or to the censoring time for those without an event) across participants.

A simple mocked-up version of this figure using simulated data is displayed below.

Questions/Comments

6.5 Tables

6.5.1 List of Tables

Table	Number	Description
Baseline	1	Compares 4-treatment groups on baseline risk factors, history of cvd, and baseline renal function
Outcome Comparison by Group	2	Treatment group differences of micro/macro vascular outcomes
Risk Factors	3	Treatment group differences of micro/macrovascular risk factors
Mediation by HbA1c level	4	Mediation of treatment group effects on micro/macro vascular outcomes by HbA1c level
Heterogeneity	5	Heterogeneity of treatment group effects by baseline risk factors
Demographics	S1	Demographic breakdown of study cohort

6.5.2 Table 1: Baseline Table

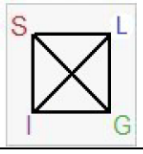
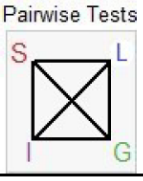
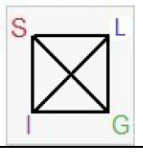
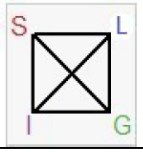
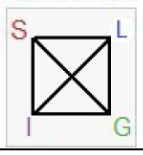
Table 1. Baseline characteristics relevant to microvascular and cardiovascular outcomes in the whole cohort and by treatment group

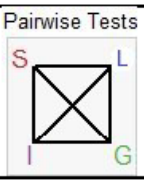
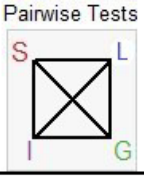
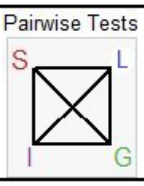
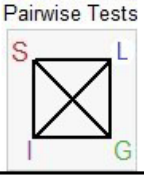
	All	Glargine	Glimepiride	Liraglutide	Sitagliptin
N	xxx	xxx	xxx	xxx	xxx
Clinical risk factors for microvascular and macrovascular disease					
BMI ($kg \cdot m^{-2}$)	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
Weight (kg)	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
Diabetes duration (years)	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
HbA1c (%)	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
HbA1c < 7%	N (%)	N (%)	N (%)	N (%)	N (%)
Lipids LDL (mg/dL) HDL (mg/dL) Triglycerides (mg/dL)	 x.x±x.x x.x±x.x x.x±x.x	 x.x±x.x x.x±x.x x.x±x.x	 x.x±x.x x.x±x.x x.x±x.x	 x.x±x.x x.x±x.x x.x±x.x	 x.x±x.x x.x±x.x x.x±x.x
Lipid-lowering medications Any Statin Other	 N (%) N (%) N (%)	 N (%) N (%) N (%)	 N (%) N (%) N (%)	 N (%) N (%) N (%)	 N (%) N (%) N (%)
Hypertension SBP > 140 mm/Hg SBP (mm/Hg) DBP (mm/Hg)	 N (%) x.x±x.x x.x±x.x	 N (%) x.x±x.x x.x±x.x	 N (%) x.x±x.x x.x±x.x	 N (%) x.x±x.x x.x±x.x	 N (%) x.x±x.x x.x±x.x
Blood pressure medications Any BP meds ACEi/ARB Other BP meds	 N (%) N (%) N (%)	 N (%) N (%) N (%)	 N (%) N (%) N (%)	 N (%) N (%) N (%)	 N (%) N (%) N (%)
Baseline Renal Function[1]					
UACR ($mg \cdot g^{-1}$)[1]	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
Microalbuminuria (UACR ≥ 30 $mg \cdot g^{-1}$)	N (%)	N (%)	N (%)	N (%)	N (%)
Macroalbuminuria (UACR ≥ 300 $mg \cdot g^{-1}$)	N (%)	N (%)	N (%)	N (%)	N (%)
eGFR ($mL \cdot min^{-1} \cdot (1.73\ m^2)^{-1}$)	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
eGFR < 60 $ml \cdot min^{-1} \cdot (1.73\ m^2)^{-1}$	N (%)	N (%)	N (%)	N (%)	N (%)
Baseline DSPN[2]	N (%)	N (%)	N (%)	N (%)	N (%)
Baseline MACE[3]	N (%)	N (%)	N (%)	N (%)	N (%)

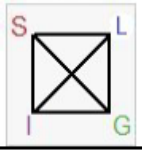
[1]UACR - urinary albumin to creatinine ratio.
[2]DSPN - distal symmetric polyneuropathy, defined using the MNSI questionnaire and foot exam
[3]MACE-Major adverse cardiovascular events.
Numbers in table are Mean ± SD, unless stated otherwise.

6.5.3 Table 2: Treatment group differences in micro/macro vascular outcomes

Table 2 Assess the treatment group differences of micro/macro vascular status

	Total (N=5047)	**Glargine (N=XXXX)**	**Glimepiride (N=XXXX)**	**Liraglutide (N=XXXX)**	**Sitigliptin (N=XXXX)**	**Pairwise Treatment Comparisons[1]**
MACE						
n (%)						
Crude rate per 100 person-years (SE)						
Pairwise hazard ratios (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 
Hazard ratio compared to all other treatments combined (SE)	--					G: p= L: p= S: p= I: p=
Pairwise RMST ratio (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 
RMST ratio compared to all other treatments (SE)	--					G: p= L: p= S: p= I: p=
CHF						
n (%)						
Crude rate per 100 person-years (SE)						
Pairwise hazard ratios (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 
Hazard ratio compared to all other treatments combined (SE)	--					G: p= L: p= S: p= I: p=
Pairwise RMST ratio (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 
RMST ratio compared to all other treatments (SE)	--					G: p= L: p= S: p= I: p=
DSPN						
n (%)						
Crude rate per 100 person-years (SE)	--					
Pairwise hazard ratios (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 

Hazard ratio compared to all other treatments combined (SE)	--					G: p= L: p= S: p= I: p=
Pairwise RMST ratio (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 
RMST ratio compared to all other treatments (SE)	--					G: p= L: p= S: p= I: p=
eGFR < 60 mL·min^{-1}·(1.73 m^2)$^{-1}$						
n (%)						
Crude rate per 100 person-years (SE)	--					
Pairwise hazard ratios (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 
Hazard ratio compared to all other treatments combined (SE)	--					G: p= L: p= S: p= I: p=
Pairwise RMST ratio (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 
RMST ratio compared to all other treatments (SE)	--					G: p= L: p= S: p= I: p=
Microalbuminuria (UACR ≥ 30 mg·g^{-1})						
n (%)						
Crude rate per 100 person-years (SE)	--					
Pairwise hazard ratios (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 
Hazard ratio compared to all other treatments combined (SE)	--					G: p= L: p= S: p= I: p=
Pairwise RMST ratio (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 
RMST ratio compared to all other treatments (SE)	--					G: p= L: p= S: p= I: p=
Macroalbuminuria (UACR ≥ 300 mg·g^{-1})						

n (%)						
Crude rate per 100 person-years (SE)	--					
Pairwise hazard ratios (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 
Hazard ratio compared to all other treatments combined (SE)	--					G: p= L: p= S: p= I: p=
Pairwise RMST ratio (SE) Glargine (SE) Glimepiride (SE) Liraglutide (SE) Sitigliptin (SE)	 -- -- -- --	 -- -- -- --	 -- -- --	 -- --	 --	Pairwise Tests 
RMST ratio compared to all other treatments (SE)	--					G: p= L: p= S: p= I: p=

* $p \leq 0.05$ (from test that hazard ratio equals 1)
** $p \leq 0.01$ (from test that hazard ratio equals 1)
*** $p \leq 0.001$ (from test that hazard ratio equals 1)

[1] Boxes in this column graphically display the results of testing pairwise treatment effects. Each corner of the box represents one of the four treatments (G = Glimepiride, L = Liraglutide, S = Sitagliptin, I=Insulin Glargine), and lines connect the treatments that differ significantly. Dotted lines indicate $p \leq 0.05$, dashed lines indicate $p \leq 0.01$, and solid lines indicate $p \leq 0.001$.

6.5.4 Table 3: Profiles (baseline, year1, year4) of micro/macrovascular risk factors by treatment groups

Table 3 Assess the treatment group differences in levels of micro/macro vascular risk factors during follow-up

Characteristic (p)[2]	**Total[1] (N=5047)**	**Glargine[1] (N=XXXX)**	**Glimepiride[1] (N=XXXX)**	**Liraglutide[1] (N=XXXX)**	**Sitigliptin[1] (N=XXXX)**
Direct measures of micro/macro[3]					
UACR(p=x.xx)					
Baseline	x.x±x.x	-	-	-	-
GRADE 1yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
GRADE 5yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
eGFR(p=x.xx)					
Baseline	x.x±x.x	-	-	-	-
GRADE 1yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
GRADE 5yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
eGFR < 60 (p=x.xx)					
Baseline	N(%)	-	-	-	-
GRADE 1yr	N(%)	N(%)	N(%)	N(%)	N(%)
GRADE 5yr	N(%)	N(%)	N(%)	N(%)	N(%)
MicroAlb (p=x.xx)					
Baseline	N(%)	-	-	-	-
GRADE 1yr	N(%)	N(%)	N(%)	N(%)	N(%)
GRADE 5yr	N(%)	N(%)	N(%)	N(%)	N(%)
MacroAlb (p=x.xx)					
Baseline	N(%)	-	-	-	-
GRADE 1yr	N(%)	N(%)	N(%)	N(%)	N(%)
GRADE 5yr	N(%)	N(%)	N(%)	N(%)	N(%)
Hypertension[3]					
SBP > 140 mm/Hg (p=x.xx)					
Baseline	N(%)	-	-	-	-
GRADE 1yr	N(%)	N(%)	N(%)	N(%)	N(%)
GRADE 5yr	N(%)	N(%)	N(%)	N(%)	N(%)
SBP(p=x.xx)					
Baseline	x.x±x.x	-	-	-	-
GRADE 1yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
GRADE 5yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
DBP(p=x.xx)					
Baseline	x.x±x.x	-	-	-	-
GRADE 1yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
GRADE 5yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
Any BP meds (p=x.xx)					
Baseline	N(%)	-	-	-	-
GRADE 1yr	N(%)	N(%)	N(%)	N(%)	N(%)
GRADE 5yr	N(%)	N(%)	N(%)	N(%)	N(%)

ACEi/ARB (p=x.xx)					
Baseline	N(%)	-	-	-	-
GRADE 1yr	N(%)	N(%)	N(%)	N(%)	N(%)
GRADE 5yr	N(%)	N(%)	N(%)	N(%)	N(%)
Dysliplidemia[3]					
HDL(p=x.xx)					
Baseline	x.x±x.x	-	-	-	-
GRADE 1yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
GRADE 5yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
LDL(p=x.xx)					
Baseline	x.x±x.x	-	-	-	-
GRADE 1yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
GRADE 5yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
Triglycerides (p=x.xx)					
Baseline	x.x±x.x	-	-	-	-
GRADE 1yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
GRADE 5yr	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x	x.x±x.x
Statin Use (p=x.xx)					
Baseline	N(%)	-	-	-	-
GRADE 1yr	N(%)	N(%)	N(%)	N(%)	N(%)
GRADE 5yr	N(%)	N(%)	N(%)	N(%)	N(%)

[1]Table estimates are mean ± SE or N(%) from a 4-group ANCOVA model for group adjusted for baseline, time since randomization and treatment group as covariates. Note, model estimates assume a common baseline value for risk factors (given in total column) in all treatment groups due to randomization.

[2]p-value from test of interaction of time since randomization and treatment group. Used to assess heterogeneity in risk factors over time by treatment

[3] Urinary Albumin-Creatinine Ratio (UACR) units = $mg \cdot g^{-1}$; estimated Glomerular Filtration Rate(eGFR) units = $mL \cdot min^{-1} \cdot (1.73\ m^2)^{-1}$; MicroAlb = UCAR ≥ 30; MacroAlb = UCAR ≥ 300; Systolic Blood Pressure(SBP); Diastolic Blood Pressure(DBP) units = mm/Hg; High Density Lipoproteins(HDL); Low Density Lipoproteins(LDL); Triglyceride units=mg/dL

6.5.5 Table 4: Mediation of treatment group effects on micro/macro vascular outcomes by HbA1c levels

Table 4. Assessment of mediation of treatment group effects by HbA1c

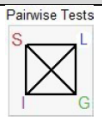
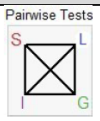
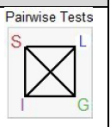
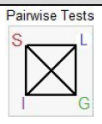
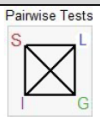
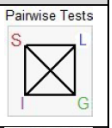
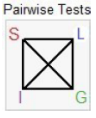
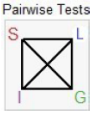
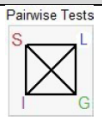
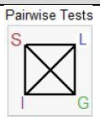
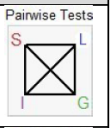
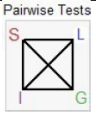
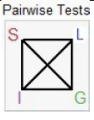
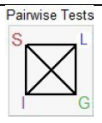
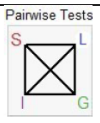
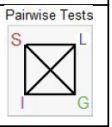
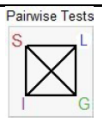
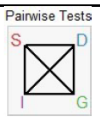
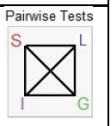
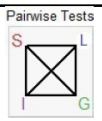
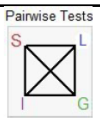
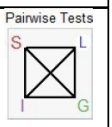
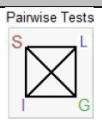
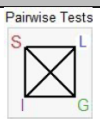
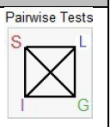
Outcome[1]	Model Adjustment	Glimepiride vs others		Liraglutide vs others		Sitagliptin vs others		Insulin Glargine vs others	
		HR (SE)[2]	% Med[1,2]	HR (SE)[2]	% Med[1,2]	HR (SE)[2]	% Med[1,2]	HR (SE)[2]	% Med[1,2]
MACE	None	θ_{01} (σ_{01})	-	θ_{02} (σ_{02})	-	θ_{03} (σ_{03})	-	θ_{04} (σ_{04})	-
	HbA1c	θ_{11} (σ_{11})	φ_{11} (σ_{11})	θ_{12} (σ_{12})	φ_{12} (σ_{12})	θ_{13} (σ_{13})	φ_{13} (σ_{13})	θ_{14} (σ_{14})	φ_{14} (σ_{14})
Heart Failure	None		-		-		-		-
	HbA1c								
DSPN	None		-		-		-		-
	HbA1c								
eGFR < 60	None		-		-		-		-
	HbA1c								
UACR ≥ 30 mg·g^{-1}	None		-		-		-		-
	HbA1c								
UACR ≥ 300 mg·g^{-1}	None		-		-		-		-
	HbA1c								

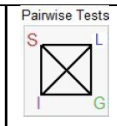
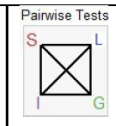
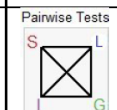
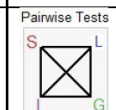
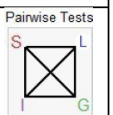
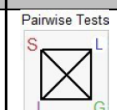
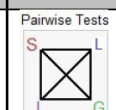
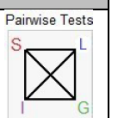
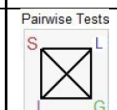
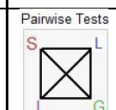
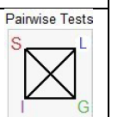
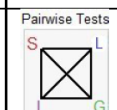
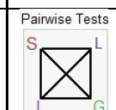
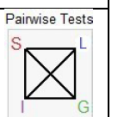
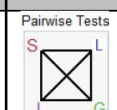
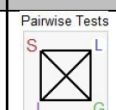
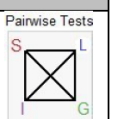
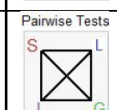
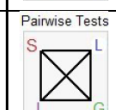
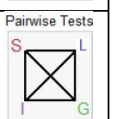
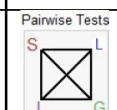
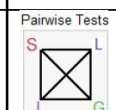
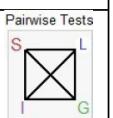
[1]Percent mediation was calculated as the relative change in the hazard ratio for treatment group in a model adjusted HbA1c relative to an unadjusted model. For example, φ_{11} is the estimate of the percent change in the hazard ratio (Glimepiride vs others) in the models adjusted for HbA1c relative to the unadjusted model and is calculated as $(\theta_{11}/\theta_{01} - 1)\cdot 100$ where the HR estimates (θ's) are calculated from the appropriate contrasts of the model coefficients from Cox proportional hazards models.

[2]An asterisk superscript indicates that the estimate is significantly different than 1 (HR) or 0 (%Med)

6.5.6 Table 5: Heterogeneity of treatment group differences of outcomes by baseline subgroups

Table 5. Heterogeneity of treatment group effects by subgroups

Subgroup (p)[1]	Treatment Group		Cardiovascular Outcome Event Rates[2]									
			MACE[3]			CHF[3]			DSPN[3]			
Total		**N**	**N_e**	**Rate(CI)**	**Tests[4]**	**N_e**	**Rate(CI)**	**Tests[4]**	**N_e**	**Rate(CI)**	**Tests[4]**
	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Gender (p=.xx)											
Male	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Female	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Race/Ethnicity (p=.xx)											
White	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Black	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Hispanic	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Asian	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Am Indian	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Age (p=.xx)											
< 45 yrs	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	

45-59	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
60+	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
HbA1c (p=.xx)											
Tertile 1 %	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Tertile 2	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Tertile 3	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Diabetes Duration (p=.xx)											
Tertile 1 yrs	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Tertile 2	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
Tertile 3	Glargine	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Glimeperide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Liraglutide	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	
	Sitagliptin	N	n	r(ll, ul)		n	r(ll, ul)		n	r(ll, ul)	

[1]Subgroup levels (p-value for test of homogeneity of treatment effect across subgroup levels)

[2]Event rates are crude rates expressed as number of events per 100 patient-years of follow-up

[3]MACE= Major adverse cardiac events; CHF=congestive heart failure; DSPN=Distal symmetric polyneuropathy

[4]Boxes in this column graphically display the results of testing pairwise treatment effects. Each corner of the box represents one of the four treatments (G = Glimepiride, L = Liraglutide, S = Sitagliptin, I=Insulin Glargine), and lines connect the treatments that differ significantly. Dotted lines indicate $p \leq 0.05$, dashed lines indicate $p \leq 0.01$, and solid lines indicate $p \leq 0.001$.

6.5.7 Table S1: Demographics of the study cohort

	All	Glargine	Glimepiride	Liraglutide	Sitagliptin
N	xxx	xxx	xxx	xxx	xxx
Demographics					
Age category					
< 45	N (%)	N (%)	N (%)	N (%)	N (%)
45-59	N (%)	N (%)	N (%)	N (%)	N (%)
60+	N (%)	N (%)	N (%)	N (%)	N (%)
Race					
White	N (%)	N (%)	N (%)	N (%)	N (%)
Black	N (%)	N (%)	N (%)	N (%)	N (%)
Hispanic	N (%)	N (%)	N (%)	N (%)	N (%)
Asian	N (%)	N (%)	N (%)	N (%)	N (%)
Am Ind	N (%)	N (%)	N (%)	N (%)	N (%)
Sex					
Female	N (%)	N (%)	N (%)	N (%)	N (%)
Male	N (%)	N (%)	N (%)	N (%)	N (%)

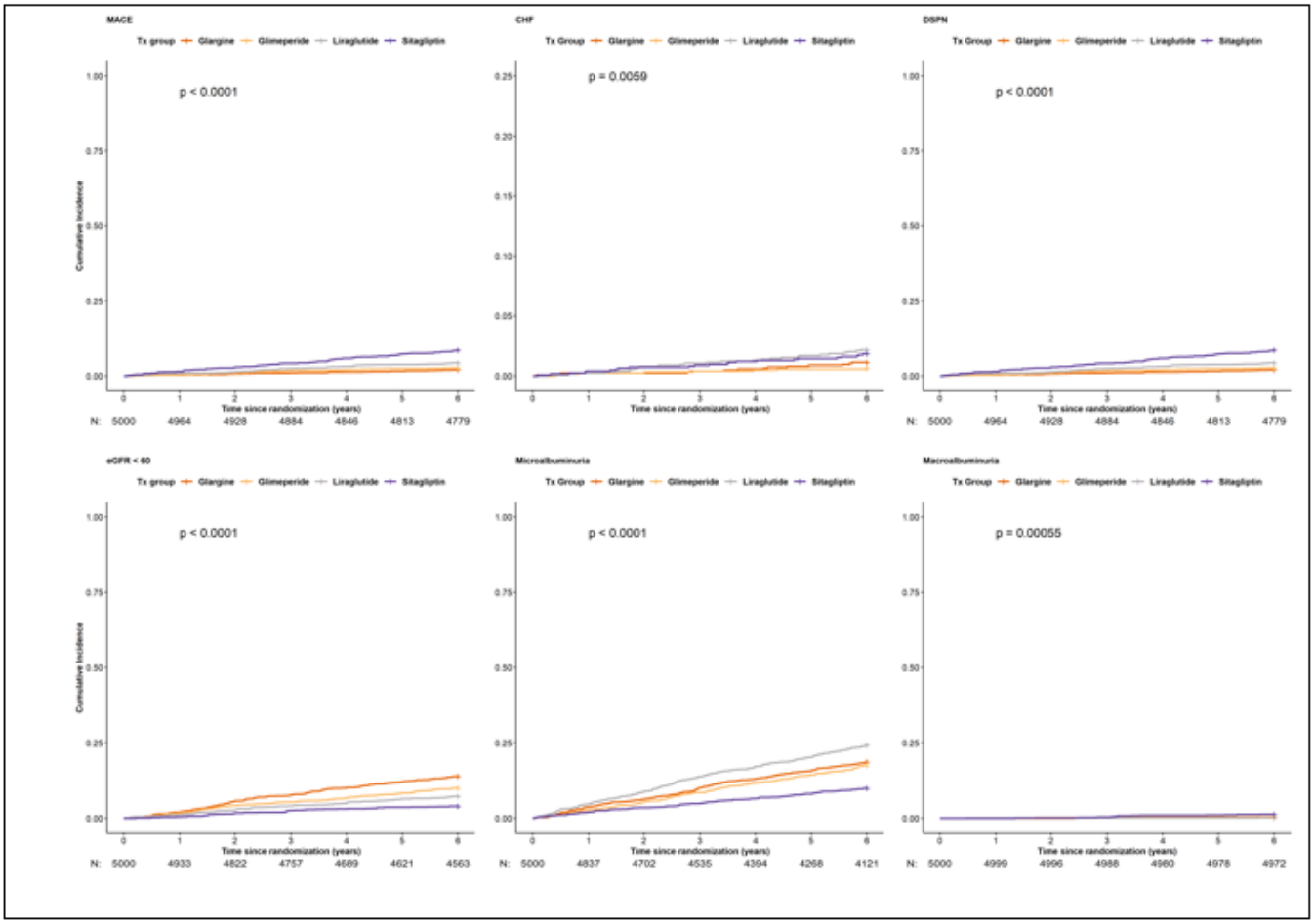
Supplementary Table #1. Demographics of the study population

6.6 Figures

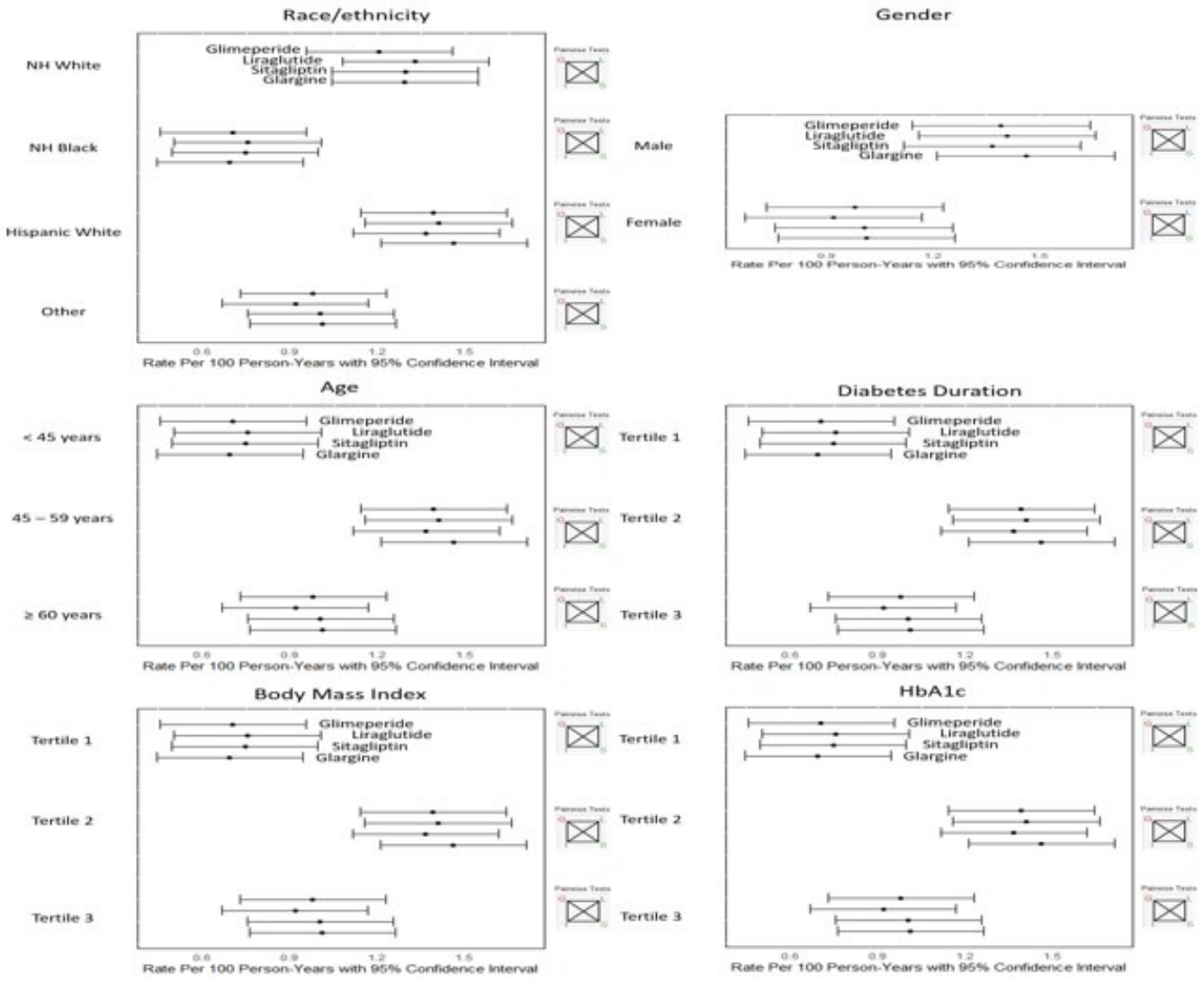
6.6.1 List of Figures

Figure	Number	Description
Kaplan-Meier	1	Cumulative incidence of micro/macro vascular outcomes by treatment group
Heterogeneity	2	Heterogeneity of treatment group effects by baseline risk factors

6.6.2 Figure 1: Cumulative incidence of micro/macro vascular outcomes by treatment group



6.6.3 Figure 2: Heterogeneity of treatment group differences of outcomes by baseline subgroups



7 STATISTICAL CONSIDERATIONS

7.1 Statistical principles and issues

7.1.1 Significance level of tests

A significance level of α=0.05 will be used for all statistical tests, unless otherwise specified. Comparisons among the treatment groups will be adjusted for the number of tests conducted, 6 for pairwise comparisons and 4 for each group versus the average of the others. Unless stated otherwise, the adjusted p-values are obtained from application of the closed testing principle (7). In cases where the closed testing adjustment cannot be readily applied, then the Holm adjustment will be employed. Otherwise, p-values will be designated as "nominal" or "simple" p-values.

7.1.2 Intention-to-treat analyses

Unless otherwise specified, all available data for all randomized participants (i.e., the full analysis set) will be included in analyses, and data will be analyzed according to the randomly assigned treatment group, regardless of adherence to assigned treatment and/or compliance with the study protocol, according to intention-to-treat principles.

7.1.3 Checking the proportional hazards assumption for the Cox proportional hazards model

For analyses based on the Cox proportional hazards model, the assumption of proportional hazards will be tested using the test of Lin (8). If the test of proportional hazards is significant (i.e., hazards are assessed to be non-proportional), then the coefficients from the Cox model will be interpreted (approximately) as average log hazard ratios, inferences (standard errors, confidence intervals, and p-values) will be based on the robust information sandwich covariance estimates (3), and the robust model score test will be used to test for treatment group differences (2).

7.1.4 Adjustments for multiple pairwise comparisons among the treatment groups

Since there are 4 treatment groups, there are 6 possible pairwise comparisons among the treatment groups. A closed testing approach will be used to account for multiple pairwise comparisons among the treatment groups (7). First, an omnibus T^2-like test will be conducted to test for any differences among the 4 treatment groups. If that test is significant at the specified significance level α, then each of the 3-group sub-hypotheses (i.e., test for differences among 3 of the treatment groups) will be tested at significance level α. Each of the pairwise comparisons can be tested at significance level α if all of the relevant higher-order hypotheses (i.e., 4-group and relevant 3-group hypotheses) are significant at significance level α. See the table below for an outline of the null hypotheses in the testing hierarchy that must be significant to allow for testing of each pairwise comparison. The hypothesis testing tree is given in the table below. Each column gives the series of null hypotheses that are tested in order to establish whether a given order 1 (i.e. 2-group) comparison is significant at level α. For example, the column with header 1 vs 2 gives the series of null hypotheses that are tested (all at level α) in order to establish whether treatment group 1 is significantly different from treatment group 2 at level α. Likewise for the 5 other pairwise group comparisons.

Pairwise Comparison	**1 vs 2**	**1 vs 3**	**1 vs 4**	**2 vs 3**	**2 vs 4**	**3 vs 4**
Order 3(4- group comp)[1]	H_0, 1234	H_0, 1234	H_0, 1234	H_0, 1234	H_0, 1234	H_0, 1234
Order 2(3- group comp)[2]	H_0, 12,13	H_0, 13,12	H_0, 14,12	H_0, 23,12	H_0, 24,12	H_0, 34,12
	H_0, 12,14	H_0, 13,14	H_0, 14,13	H_0, 23,13	H_0, 24,13	H_0, 34,13
	H_0, 12,23	H_0, 13,23	H_0, 14,23	H_0, 23,14	H_0, 24,14	H_0, 34,14

	$H_0, 12{,}24$	$H_0, 13{,}24$	$H_0, 14{,}24$	$H_0, 23{,}24$	$H_0, 24{,}23$	$H_0, 34{,}23$
	$H_0, 12{,}34$	$H_0, 13{,}34$	$H_0, 14{,}34$	$H_0, 23{,}34$	$H_0, 24{,}34$	$H_0, 34{,}24$
Order 1(2- group comp)[3]	$H_0, 12$	$H_0, 13$	$H_0, 14$	$H_0, 23$	$H_0, 24$	$H_0, 34$

[1] $H_0, 1234$ is the null hypothesis that $\mu_1 = \mu_2 = \mu_3 = \mu_4$. This is a 3 df test.

[2] H_0, ab, cd is the null hypothesis that $\mu_a = \mu_b$ and $\mu_c = \mu_d$. These are 2 df tests.

[3] H_0, ab is the null hypothesis that $\mu_a = \mu_b$. These are 1 df tests.

where a,b,c,d ∈ {1,2,3,4}

Note that some of the order 2 tests are equivalent (e.g. $H_0, 13{,}12 \equiv H_0, 13{,}23$).

7.1.5 Comparing each treatment to all other treatments combined

There is interest in testing whether the effect of each treatment differs from the other 3 treatment groups combined. Let θ_k be the log(hazard ratio) comparing the hazard for treatment group k=1,2,3 to the hazard for reference treatment group k=4. For each treatment group, we would test the null hypothesis that the average of the estimated hazard ratios comparing each of the other treatments to the treatment of interest equals 1. In other words, we would test each of the following 4 null hypotheses (i.e., one hypothesis per treatment group):

$H_{01}: e^{\theta_2-\theta_1} + e^{\theta_3-\theta_1} + e^{-\theta_1} = 3$

$H_{02}: e^{\theta_1-\theta_2} + e^{\theta_3-\theta_2} + e^{-\theta_2} = 3$

$H_{03}: e^{\theta_1-\theta_2} + e^{\theta_2-\theta_3} + e^{-\theta_3} = 3$

$H_{04}: e^{\theta_1} + e^{\theta_2} + e^{\theta_3} = 3$

A closed testing approach will be used to account for multiple comparisons, according to the procedure described in (4). The closed testing hierarchy would start with the 3-df test of the joint hypothesis $\theta_1 = \theta_2 = \theta_3 = 0$. The next stage of the closed testing hierarchy would be to test the intersections of the elementary hypotheses listed above (e.g., $H_{01} \cap H_{02}$). The last stage would be to test the elementary hypotheses listed above. For example, the elementary hypothesis H_{01} would be rejected at significance level α if $H_{01}, H_{01} \cap H_{02}, H_{01} \cap H_{03}, H_{01} \cap H_{04}$, and the joint hypothesis $\theta_1 = \theta_2 = \theta_3 = 0$ are all significant at significance level α.

7.1.6 Adjustments for multiple comparisons for subgroup analyses

One of the objectives of this paper is to assess treatment group differences within baseline subgroups (e.g., tertiles of HbA1c). There are 6 possible pairwise comparisons among the treatment groups within each subgroup. A closed testing approach will also be used to account for multiple comparisons for testing treatment group differences within subgroups (7). Here, we describe the general closed testing approach for the case with all 4 treatment groups and 3 subgroups (e.g., tertiles of HbA1c), where θ_{jk} is the measure of treatment difference between treatment k=1,2,3 and the reference treatment k=4 within subgroup j=a,b,c. First, an overall test of the null hypothesis of homogeneity of treatment effects across all subgroups would be tested:

$$\begin{aligned} H_0, abc: \theta_{a1} &= \theta_{b1} = \theta_{c1} \\ \theta_{a2} &= \theta_{b2} = \theta_{c2} \\ \theta_{a3} &= \theta_{b3} = \theta_{c3} \end{aligned}$$

If this test is significant at the specified significance level (α=0.05), then tests of null hypotheses of homogeneity of treatment effects between pairs of subgroups would be tested:

$H_{0,ab}: \theta_{a1} = \theta_{b1}, \theta_{a2} = \theta_{b2}, \theta_{a3} = \theta_{b3}$

$H_{0,ac}: \theta_{a1} = \theta_{c1}, \theta_{a2} = \theta_{c2}, \theta_{a3} = \theta_{c3}$

$H_{0,bc}: \theta_{b1=}\theta_{c1}, \theta_{b2} = \theta_{c2}, \theta_{b3} = \theta_{c3}$

Then if any two of these tests were significant at the specified significance level (α=0.05), then within the intersection subgroup, tests of pairwise treatment comparisons can proceed in a similar manner as described in the previous section (related to adjustment of multiple pairwise comparisons among treatment groups). For example, if the tests of $H_{0,ab}$ and $H_{0,ac}$ were both significant, then testing of pairwise treatment comparisons can proceed within subgroup a.

7.1.7 Calculation of confidence intervals adjusted for multiple comparisons

For analyses with multiple comparisons (e.g., pairwise treatment comparisons, comparisons of each treatment group vs. all others combined, subgroup analyses), confidence intervals for effect estimates will be calculated based on a method that controls the family-wise type 1 error for multiple comparisons.

8 DISCUSSION POINTS FOR WRITING GROUP

8.1 Considerations

Listing of scientific/statistical issues that need to be discussed by the writing group. For example, how to handle cases where a participant did not have a final lab-based outcome measure because they experienced a clinical event that is directly related to the lab-based outcome (e.g. did not have an IVGTT because was diagnosed with diabetes prior to the IVGTT visit)

8.2 Limitations

Discuss any limitations of the study (e.g. no baseline measure of outcome, potential selection biases in the study sample etc)

9 APPENDIX A: Dataset Request

9.1 Table of Variables

Measure	Variable	units	Assessment Visits	Notes
Treatment	assign		Baseline	
Age at randomization	age	(yrs)	Baseline	< 45 45-59 60+
Gender	gender	MF.	Baseline	M/F
Race	race	Race.	Baseline	White Black Hispanic Asian Am Indian
Weight	weight	(kg)	Baseline quarterly	

BMI	bmi	(kg/m2)	Baseline quarterly	Categories 22 - < 30 30 - < 35 >=35
SBP	sbp	(mmHg)	Baseline quarterly	
DBP	dbp	(mmHg)	Baseline quarterly	
Hypertension	hyper		Baseline quarterly	SBP > 140
Any BP meds	anybp	YN.	Baseline quarterly	
ACEi/ARB	aceiarb	YN.	Baseline quarterly	
Other BP meds	otherbp	YN.	Baseline quarterly	
HDL	hdl	mg/dL	Baseline annual	
LDL	ldl	mg/dL	Baseline annual	
Triglycerides	trig	mg/dL	Baseline annual	
Any Lipid meds	anyllm	YN.	Baseline quarterly	
Statin meds	statins	YN.	Baseline quarterly	
Other Lipid meds	othllm	YN.	Baseline quarterly	
HbA1c	hba1c	mg/dL	Baseline quarterly	
Diabetes Duration	diabdur	yrs	Baseline quarterly	Time from diagnosis to visit
UACR	acr	mg/g	Baseline semi-annual	
UACR ≥ 30	microalb	mg/g	Baseline semi-annual	confirmed?
UACR ≥ 300	macroalb	mg/g	Baseline semi-annual	confirmed?
eGFR	egfr	mL/min/ 1.73m2	Baseline Annual	
eGFR < 60	egfr	mL/min/ 1.73m2	Baseline Annual	confirmed?
MACE	mace	YN.	Event Time	adjudicated ?
CHF	chf	YN.	Event Time	adjudicated ?
DSPN	DSPN	YN.	Baseline Annual	EDIC definition (1)

YN. = Yes or No format (0=No; 1=Yes) MF. = Male or Female format (0=Female; 1=Male)

REFERENCES

1. Herman WH, Pop-Busui R, Braffett BH, et al. Use of the michigan neuropathy screening instrument as a measure of distal symmetrical peripheral neuropathy in type 1 diabetes: Results from the diabetes control and complications trial/epidemiology of diabetes interventions and complications. *Diabetic medicine*. 2012;29(7):937–944.

2. Lachin JM. Biostatistical methods: The assessment of relative risks. 2nd ed. New York: John Wiley & Sons; 2011.

3. Lin DY, Wei LJ. The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association*. 1989;84(408):1074–1078.

4. Lachin JM, Bebu I. Closed testing of each group versus the others combined in a multiple group analysis. *Clinical Trials*. 2020;17(1):77–86.

5. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*. 1986;51(6):1173–1182.

6. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*. 1984;71(3).

7. Lachin JM, Bebu I, Larsen MD, et al. Closed testing using surrogate hypotheses with restricted alternatives. *PLoS ONE*. 2019;(7):1–18.

8. Lin DY. Goodness-of-fit analysis for the cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association*. 1991;86(415):725–728.

April 8, 2022

Amended Final Statistical Analysis Plan as of April 8, 2022

Treatment group differences in micro/macrovascular outcomes among four initial treatments added to metformin in early type 2 diabetes (OP2)

Table of Contents

1 GENERAL INFORMATION

This document is an amended final statistical analysis plan. This document was prepared and reviewed by John M. Lachin, Naji Younes, Heidi Krause-Steinrauf and Mark Tripputi.

GRADE paper number	OP 2
Analysis Category	End of Study
Writing Group Chairs	David Nathan, John Lachin
Writing Group Members	David Nathan, John Lachin, John Buse, Henry Burch, Steven Kahn, Andrea Cherrington, Stephen Fortmann, Jennifer Green, Sue Kirkman, Heidi Krause-Steinrauf, Mary Larkin, Lawrence Phillips, Rodica Pop-Busui, Michael Steffes, Margaret Tiktin, Deborah Wexler, Ionut Bebu, Mark Tripputi, Naji Younes
Target Journal	NEJM
Lead Statisticians	John M. Lachin, Naji Younes

2 APPROVALS

No signatures are provided.

3 REVISION HISTORY

Version No.	Implemented by	Date
1 Initial SAP	Mark Tripputi	05/12/2021
2 Final SAP for Initial Submission	Mark Tripputi	01/19/2022
3 Final SAP for Re-submission	Mark Tripputi	04/08/2022

4 ABBREVIATIONS AND ACRONYMS

Abbreviation	Meaning
GLP-1	Glucagon Like Peptide 1
DPP-4	Dipeptidyl peptidase 4
SGLT-2	sodium-glucose cotransporter 2
MACE	Major adverse cardiovascular events
HF	Heart Failure requiring hospitalization
DPN	Diabetic Peripheral Neuropathy
eGFR	estimated glomerular filtration rate
UACR	Urine Albumin-to-Creatinine Ratio

5 STUDY OBJECTIVES

5.1 Background and justification

Type 2 diabetes (T2DM) affects more than 30 million persons in the United States, with an incidence of 1.5 million new cases per year, and more than 400 million persons world-wide. The major human and economic costs associated with T2DM are related primarily to the development of long-term diabetes-specific complications, including retinopathy, nephropathy, and neuropathy, and a 2-5 fold increased risk of non-specific cardiovascular disease (CVD). These long-term complications have been shown to be ameliorated in part by interventions that reduce chronic glycemia, as measured by glycated hemoglobin levels (HbA1c), and a target range of less than 7% (53 mmol/mol) has been established by consensus for most patients with T2DM. The estimated annual cost of diabetes in the US in 2017 was approximately $327 billion dollars per year with an increasing fraction attributed to the cost of glucose-lowering medications.

Virtually all recommendations for the management of type 2 diabetes have included metformin as the first medication to be used. Unfortunately, choosing the second medication from the ever expanding list of glucose-lowering medications to add to metformin when monotherapy fails to achieve or maintain goal glycemia is problematic owing to the dearth of any long-term head-to-head comparator studies. The purpose of the Glycemia Reduction Approaches in Type 2 Diabetes: A Comparative Effectiveness (GRADE) Study was to examine the relative effectiveness of the four most commonly used glucose-lowering medications added to metformin to maintain goal glycemia. In this paper, we report the difference between the treatment groups in the incidence of micro- and macrovascular outcomes and their risk factors. The accompanying paper reports treatment group differences in the metabolic outcomes.

5.2 Scientific objectives

1. Compare cumulative incidence of outcomes by treatment group to assess whether one or more of the treatment approaches had increased (or decreased) benefit compared with the others

2. Compare the relative efficacy of treatment groups (hazard ratios) on the risk of the micro/macro vascular outcomes. This will include both pairwise comparisons, and comparisons of each group vs the other groups combined.

3. Compare profiles (baseline, year1, year4) of micro/macrovascular risk factors by treatment groups

4. Compare differences in treatment group effects on micro/macrovascular outcomes by levels of relevant subgroup variables

6 STATISTICAL METHODS AND DATASETS

6.1 Analysis Data Set Inclusion Criteria

The study will include all GRADE randomized participants.

6.2 Outcomes to be Assessed

There are nine micro/macrovascular outcomes of primary interest in this paper:

1. Major Adverse Cardiovascular Events (MACE): consists of non-fatal MI, stroke, or cv death
2. Heart Failure (HF)
3. Any cardiovascular disease (any CVD)
4. Diabetic Peripheral Neuropathy (DPN)
5. eGFR < 60 $mL \cdot min^{-1} \cdot (1.73m^2)^{-1}$
6. Confirmed Microalbuminuria (UACR ≥ 30 $mg \cdot g^{-1}$ on 2 successive evaluations)
7. Macroalbuminuria (UACR ≥ 300 $mg \cdot g^{-1}$)
8. Cardiovascular mortality (CV mortality)
9. Total mortality

Definition of Major Adverse Cardiac Events(MACE)

- Defined as occurrence of one or more of the 3 MACE components: non-fatal MI, non-fatal stroke, cardiovascular death including fatal stroke or fatal MI
- Requires adjudication (Cardiovascular Event (MACE and Non-MACE) Adjudication Form)

Definition of Heart Failure(HF)

- Defined as diagnosis of HF requiring hospitalization with adjudication (Cardiovascular Event (MACE and Non-MACE) Adjudication Form)

Definition of any cardiovascular disease (any CVD)

- Defined as MACE (as defined above), OR unstable angina requiring hospitalization or revascularization, OR heart failure (as defined above), OR any revascularization event.

Definition of DPN

- DPN will be assessed using the MNSI questionnaire and neuropathy exam components (1)
- MNSI questionnaire (collected annually)
  - 15 items with each question scored 0 if symptom absent, 1 if present
- Neuropathy exam (annual form) consists of 5 components assessing both left and right feet
  - Appearance (Normal/Abnormal)
  - Ulceration (Y/N)
  - Ankle reflexes (Present/Absent)
  - Vibration perception using tuning fork on top of great toes (absent, reduced, present)
  - Detection of 10g monofilament on top of great toes, with 10 trials on each toe (absent, reduced 1-7, present ≥ 8)
- DPN assessments are conducted at baseline and annually during follow-up

The prevalence and incidence of DPN will be defined on the basis of these assessments as follows:

- Incident or prevalent DPN will be defined as an MNSI symptom score of ≥ 7.5 AND/OR MNSI exam score (for appearance, ulcer, reflex, vibration – excluding the monofilament) of ≥ 2.5, occurring among any participant who did not satisfy either of those criteria at baseline.

In addition, a report of any ulcerations or amputations of any part of foot or lower extremity are captured as SAEs.

Definition of eGFR < 60 outcome

- (eGFR < 60 $mL \cdot min^{-1} \cdot (1.73m^2)^{-1}$ and baseline value ≥ 60) OR ESRD death OR dialysis/transplant
- The eGFR measurement is based on the serum creatinine that is collected annually
- Dialysis OR Transplant reported as single outcome (quarterly and annual forms)
- ESRD death is adjudicated

Definition of Confirmed Microalbuminuria

- (UACR ≥ 30 $mg \cdot g^{-1}$ AND baseline value < 30) OR ESRD death OR dialysis/transplant
- UACR (collected every 6 months) threshold confirmed at two consecutive visits
- See section 7.1.8 for a description of the confirmation process for microalbuminuria

Definition of Macroalbuminuria

- Protocol defined as occurrence of UACR ≥ 300 $mg \cdot g^{-1}$ and baseline value < 300 OR ESRD death OR dialysis/transplant

Definition of CV mortality

- Death whose immediate or underlying cause was any of the following: sudden death with evidence of CVD, MACE death, or undetermined death. This is an adjudicated outcome.

Definition of Total mortality

- All-cause death

6.3 Assessment of Study Power

The GRADE protocol contains power calculations for MACE and confirmed microalbuminuria:

- With a projected incidence rate of 0.04/yr for *confirmed microalbuminuria* (indicates rate in group with highest incidence), GRADE would have 88% power to detect a 33% difference in risk between any of the 6 pairwise treatment group comparisons
- With a projected incidence rate of 0.01/yr for *MACE* (indicates rate in group with highest incidence), GRADE would have 80% power to detect a 50% difference in risk between any of the 6 pairwise treatment group comparisons. Smaller differences can be detected if each treatment group is compared with the other 3 combined treatment groups (80% power to detect a 42% difference in risk).

Event Counts

Event rates calculated using the **`grCore:/RC_3.0 events`** dataset (not the complete final dataset)

- Data for nephropathy events
  - table gives both unconfirmed (N) and confirmed counts (N confirmed) of participants with one or more microalbuminuria events. No other outcomes are confirmed.
  - See section 7.1.8 for a description of the confirmation process for microalbuminuria
- Data for MACE and HF events
  - table gives the count of participants with one or more adjudicated MACE events (cv death OR non-fatal stroke OR non-fatal MI), the component events, and HF events

The numbers of events have not been updated to reflect the final data closeout.

Event	N	N confirmed[1]	At Risk yrs	Annual Rate
Micro-albuminuria	1320	641	3.50, 4.05	0.075, 0.031
Macro-albuminuria	250	N/a	4.72	0.010
eGFR < 60	608	N/a	4.72	0.025
MACE	190	N/a	4.95	0.008
CV death	17	N/a	2.58	0.001
stroke	81	N/a	5.26	0.003
MI	107	N/a	4.94	0.004
DPN	1722	N/a	5.22	0.065
HF	103	N/a	3.11	0.0007

[1]Confirmation applies only for microalbuminuria, and not macroalbuminuria or eGFR < 60
N/a: not applicable

6.4 Statistical Analyses

All figures and tables referenced in this section are shown in sections below (Proposed Tables, Proposed Figures).

6.4.0 Treatment effect on cumulative incidence of risk factors for micro/macrovascular outcomes

Objective Compare cumulative incidence of risk factors for micro/macrovascular outcomes by treatment group (Figure 1)

Statistical Analyses

Figure 1 description (see mockup below)

A 1 x 2 panel figure.

The two panels show the cumulative incidence for hypertension (left panel) and hyperlipidemia (right panel), important risk factors for macro and microvascular outcomes.

Hypertension is defined as at least one of the following: a) History of hypertension at baseline, b) diagnosis of hypertension during GRADE, c) use of hypertensive medications regardless of reason or d) sbp>=140 or dbp>=90 on two occasions. Confirmation of sbp>=140 or dbp>=90 is not required at baseline (i.e. baseline prevalence is defined from a single elevation of either sbp or dbp)

Hyperlipidemia is defined as at least one of the following: a) Being on a lipid medication, b) Having a history of or diagnosis of hyperlipidemia or c) Having any one of the following: LDL ≥ 100 mg/dL, Triglycerides ≥150 mg/dL or HDL < 40 mg/dL for men, < 50 mg/dL for women.

6.4.1 Scientific objective 1: Treatment effect on Cumulative Incidence of outcomes

Objective Compare cumulative incidence of outcomes by treatment group (Figure 2 and 3)

Statistical Analyses

The cumulative incidence by treatment group will be estimated using a Kaplan-Meier estimator. The total number at risk at each year will be provided below each panel. The time axis will represent the time since GRADE randomization. The maximum value for the time axis is 6.5 years. The unadjusted log-rank test will be used to assess differences between the treatment groups(2).

Figure 2 description (see mockup below)

A 2 x 2 panel figure.

The 4-panel plot shows the cumulative incidence by treatment group for the 4 microvascular outcomes: moderately increased albuminuria, severely increased albuminuria, eGFR < 60 and Diabetic Peripheral Neuropathy.

Figure 3 description (see mockup below)

A 5-panel plot showing the cumulative incidence by treatment group for the CVD and mortality outcomes (any CVD, MACE, hospitalized heart failure, CV death, all-cause death).

6.4.2 Scientific objective 2: Treatment effect on hazard ratios

Objective Compare risk of outcomes between treatment groups (Table 2)

Statistical Analyses

For this table, the following statistics will be calculated for the micro/macrovascular outcomes, both overall and stratified by treatment group:

- The number of events and percent of the GRADE cohort with the outcome.
- Crude rate per 100 person-years (SE). This will be calculated as 100*(observed number of events)/(total time at risk), where the total time at risk is the sum of the time since randomization to the event (or to the censoring time for those without an event) across participants.
- Pairwise hazard ratios (SE). A Cox proportional hazards model will be fit for the outcome with treatment group as a predictor. For the purposes of this Cox model, the event times and censoring times will be calculated as time since randomization to the event or censoring, respectively. Hazard ratios and standard errors for each pairwise comparison of the treatment groups will be estimated from the Cox model. All Wald-type tests, standard errors and confidence intervals will be estimated using the robust (3) information sandwich estimator to ensure valid inferences even if the proportional hazards assumption does not apply. A joint test for differences in the hazards among any of the treatment groups will be conducted. If that joint test is significant, then pairwise log-rank tests will be conducted to test for all pairwise differences. There are a total of 6 possible pairwise comparisons among the 4 treatment groups, and therefore these tests will be adjusted for multiple comparisons using a closed testing procedure (see details in the Other statistical issues section of this document). If the joint test for differences among any of the treatment groups is significant, then the results from the pairwise testing will be indicated via footnotes to the table.
- Hazard ratio compared to all other treatments combined (SE) (4). A Cox proportional hazards model will be fit for the outcome with treatment group as a predictor. For the purposes of this Cox model, the event times and censoring times will be calculated as time since randomization to the

event or censoring respectively. For a given treatment group, the hazard ratio compared to all other treatments combined will be estimated as the average of the estimated hazard ratios comparing each of the other treatments to the given treatment group. Since there are 4 treatment groups, there would be a total of 4 tests comparing each treatment to all others combined, and therefore these tests will be adjusted for multiple comparisons using the previously described closed testing procedure (see details in the "Other statistical issues" section at the end of this document). As for the pairwise results, any significant differences will be reported as footnotes to the table

6.4.3 Scientific objective 3: Risk factor profiles by treatment group

Objective Compare profiles (baseline, year1, year4) of micro/macrovascular risk factors by treatment groups (Table 3)

Risk factors of interest include UACR, eGFR, systolic blood pressure (SBP), diastolic blood pressure (DBP), HDL, LDL, triglycerides, and prevalence of eGFR < 60, micro/macro albuminuria, SBP ≥ 140, use of blood pressure lowering medications, use of ACEi/ARBs, and use of statins.

Statistical Analyses

- GEE models will be used to estimate time averaged values of the risk factor with an unstructured covariance matrix for the multiple time point measurements. Inclusion of a time by treatment interaction terms will allow for testing of heterogeneity of time effects by treatment group. The models will be adjusted for baseline levels of the risk factor. The following estimates will be examined:
  - mean values and standard errors of risk factors overall at baseline, 1yr and 4yr post-randomization
  - mean values and standard errors of risk factors in treatment groups at 1yr and 4yrs (assumed common baseline value from overall estimate)
  - the p-value from a test of interaction of time by treatment group will be used to assess if there is heterogeneity in the time trend of estimates by treatment group.
- Any significant treatment group differences in time trends of risk factors will be reported in the text

6.4.5 Scientific objective 5: Subgroup Analyses

Assess if there are differences in treatment group effects on micro/macrovascular outcomes by levels of relevant subgroup variables (Table 5, Figure 2)

Statistical Analyses

Subgroup analyses of the treatment effects for the micro/macrovascular outcomes within subgroups based on the following baseline variables:

- age (< 45, 45-59, 60+)
- sex (male, female)
- race (White, Black, All Others)
- Ethnicity (Hispanic, Non Hispanic)
- BMI (baseline tertiles)
- HbA1c (baseline tertiles)
- diabetes duration in years (baseline tertiles)

Type I error rates will be protected using a Holm procedure (see section on statistical considerations).

Cases of significant heterogeneity (adjusted p-value < 0.05) will be reported in the text.

Questions/Comments

6.5 Tables

The shell tables presented below are copies of the tables in the manuscript submission with the data removed.

6.5.1 List of Tables

Table	Number	Description
Outcome Comparison by Group	1	ITT treatment group differences of micro/macro vascular outcomes
Outcome Comparison by Group	2	ITT treatment group differences of CVD/mortality outcomes
Baseline	S1	Compares 4-treatment groups on baseline risk factors, history of cvd, and baseline renal function
Outcome Comparison by Group	S2	Per-protocol treatment group differences of micro/macro vascular outcomes
Outcome Comparison by Group	S3	Per-protocol treatment group differences of CVD/mortality outcomes

6.5.2 Table 1: ITT analyses of microvascular outcomes

	Glargine (N=xxxx)	**Glimepiride (N=xxxx)**	**Liraglutide (N=xxxx)**	**Sitagliptin (N=xxxx)**	**Total (N=xxxx)**
Moderately increased albuminuria (p=)[1]					
N(%)					
Crude Rate per 100 patient years					
Pairwise HR (SE)					
Glargine	---				
Glimepiride	---	---			
Liraglutide	---	---	---		

Sitagliptin	---	---	---	---	
One vs. others combined HR (SE)					
Severely increased albuminuria (p=)[1]					
N(%)					
Crude Rate per 100 patient years					
Pairwise HR (SE)					
Glargine	---				
Glimepiride	---	---			
Liraglutide	---	---	---		
Sitagliptin	---	---	---	---	
One vs. others combined HR (SE)					
eGFR < 60 ml/min/m^2 (p=)[1]					
N(%)					
Crude Rate per 100 patient years					
Pairwise HR (SE)					
Glargine	---				
Glimepiride	---	---			
Liraglutide	---	---	---		
Sitagliptin	---	---	---	---	
One vs. others combined HR (SE)					
DPN(p=)[1]					
N(%)					
Crude Rate per 100 patient years					
Pairwise HR (SE)					
Glargine	---				
Glimepiride	---	---			
Liraglutide	---	---	---		
Sitagliptin	---	---	---	---	
One vs. others combined HR (SE)					

[1]From a joint test for **differences in** the hazards among any of the 4 **treatment groups**, based on a Cox proportional hazards model with treatment group as the only predictor variable.

Glargine- Insulin glargine 100u/mL.

Moderately increased albuminuria -urine albumin creatinine ratio ≥30 mg/gm, confirmed.

Severely increased albuminuria -urine albumin creatinine ratio >300 mg/gm.

eGFR- estimated glomerular filtration rate.

Participants who developed incident end-stage kidney disease (dialysis, transplantation or kidney disease mortality) during the study were included in each albuminuria outcome.

6.5.3 Table 2: ITT analyses of cardiovascular and mortality outcomes.

	Glargine (N=xxxx)	Glimepiride (N=xxxx)	Liraglutide (N=xxxx)	Sitagliptin (N=xxxx)	Total (N=xxxx)
Any CVD[1] (p=)[2]					
N(%)					
Crude Rate per 100 patient years					
Pairwise HR (SE)					
Glargine	---				
Glimepiride	---	---			
Liraglutide	---	---	---		
Sitagliptin	---	---	---	---	
One vs. others combined HR (SE)					
MACE[3] (p=)[2]					
N(%)					
Crude Rate per 100 patient years					
Pairwise HR (SE)					
Glargine	---				
Glimepiride	---	---			
Liraglutide	---	---	---		
Sitagliptin	---	---	---	---	
One vs. others combined HR (SE)					
Heart failure[4] (p=)[2]					
N(%)					
Crude Rate per 100 patient years					
Pairwise HR (SE)					
Glargine	---				
Glimepiride	---	---			
Liraglutide	---	---	---		
Sitagliptin	---	---	---	---	
One vs. others combined HR (SE)					
Cardiovascular death (p=)[2]					

N(%)				
Crude Rate per 100 patient years				
Pairwise HR (SE)				
Glargine	---			
Glimepiride	---	---		
Liraglutide	---	---	---	
Sitagliptin	---	---	---	---
One vs. others combined HR (SE)				
All deaths (p=)[2]				
N(%)				
Crude Rate per 100 patient years				
Pairwise HR (SE)				
Glargine	---			
Glimepiride	---	---		
Liraglutide	---	---	---	
Sitagliptin	---	---	---	---
One vs. others combined HR (SE)				

[1]Any CVD- first of any MACE, unstable angina requiring hospitalization or revascularization, heart failure requiring hospitalization, or any revascularization event.

[2]From a joint test for differences in the hazards among any of the 4 treatment groups, based on a Cox proportional hazards model with treatment group as the only predictor variable

[3]MACE- major adverse cardiovascular events including CVD death, non-fatal myocardial infarction or non-fatal stroke.

[4]Hospitalized heart failure.

Glargine- Insulin glargine 100u/mL

6.5.4 Table S1: Demographics of the study cohort

Characteristic	Total	Glargine*	Glimepiride	Liraglutide	Sitagliptin
N participants	N	N	N	N	N
Clinical Risk Factors					
BMI (kg/m^2)	mean ± SE	mean ± SE	mean ± SE	mean ± SE	mean ± SE
Weight (kg)	mean ± SE	mean ± SE	mean ± SE	mean ± SE	mean ± SE
Diabetes Duration (years)	mean ± SE	mean ± SE	mean ± SE	mean ± SE	mean ± SE
HbA1c (%)	mean ± SE	mean ± SE	mean ± SE	mean ± SE	mean ± SE
HbA1c <7%	N(%)	N(%)	N(%)	N(%)	N(%)
Lipids					
LDL (mg/dL)	mean ± SE	mean ± SE	mean ± SE	mean ± SE	mean ± SE
HDL (mg/dL)	mean ± SE	mean ± SE	mean ± SE	mean ± SE	mean ± SE
Triglycerides (mg/dL)	mean ± SE	mean ± SE	mean ± SE	mean ± SE	mean ± SE
Dyslipidemia[2]	N(%)	N(%)	N(%)	N(%)	N(%)
Lipid lowering medication use					
Lipid Lowering Use (any)	N(%)	N(%)	N(%)	N(%)	N(%)
Statin Use	N(%)	N(%)	N(%)	N(%)	N(%)
Hypertension					
SBP >140 mm/Hg	N(%)	N(%)	N(%)	N(%)	N(%)
SBP (mm/Hg)	mean ± SE	mean ± SE	mean ± SE	mean ± SE	mean ± SE
DBP (mm/Hg)	mean ± SE	mean ± SE	mean ± SE	mean ± SE	mean ± SE
Hypertension[3]	N(%)	N(%)	N(%)	N(%)	N(%)
Blood pressure medication use					
BP meds (any)	N(%)	N(%)	N(%)	N(%)	N(%)
ACEi/ARB	N(%)	N(%)	N(%)	N(%)	N(%)
BP meds (other than ACEi/ARB)	N(%)	N(%)	N(%)	N(%)	N(%)
Baseline renal function					
UACR (mg/g)	mean ± SE	mean ± SE	mean ± SE	mean ± SE	mean ± SE
Moderately increased albuminuria (UACR ≥30 mg/g)	N(%)	N(%)	N(%)	N(%)	N(%)

Severely increased albuminuria (UACR >300 mg/g)	N(%)	N(%)	N(%)	N(%)	N(%)
eGFR (mL/min/1.73m^2)	mean ± SE	mean ± SE	mean ± SE	mean ± SE	mean ± SE
eGFR <60 mL/min/1.73m^2	N(%)	N(%)	N(%)	N(%)	N(%)
Baseline prevalence of DPN, stroke/MI					
DPN	N(%)	N(%)	N(%)	N(%)	N(%)
Stroke/MI[4]	N(%)	N(%)	N(%)	N(%)	N(%)

[1] Statistics are mean ± SD for continuous and N(%) for categorical characteristics

[2]At least one of the following: Taking lipid-lowering medication; history or diagnosis of dyslipidemia or hyperlipidemia; or study-measured LDL ≥ 100 mg/dL, Triglycerides 150 mg/dL or HDL <40 mg/dL for men, <50 mg/dL for women

[3]At least one of the following: Taking hypertensive medication at screening or baseline visit; history or diagnosis of hypertension at screening or baseline visit; or study-measured SBP ≥140 mmHg, or DBP ≥90 mmHg at screening or baseline visit

[4]Occurred > 1 year before randomization.

*Glargine- insulin glargine 100 U/mL

6.5.5 Table S2: Comparison of characteristics in the GRADE and NHANES populations

	GRADE	NHANES
Primary study aim	Glycemic durability of second diabetes medication after metformin	Subsample of NHANES participants meeting similar criteria (below)
Study Characteristics		
Key eligibility criteria	• Age ≥30 years • T2DM < 10 years • HbA1c 6.8-8.5% (51-69 mmol/mol) taking metformin monotherapy	• Age ≥30 years • T2DM < 10 years • HbA1c 6.8-8.5% (51-69 mmol/mol) taking metformin monotherapy
Randomized intervention	Medications representing four classes: Sulfonylurea (glimepiride), DPP-4 inhibitor (sitagliptin), GLP-1 analog (liraglutide), or insulin (glargine)	n/a
Primary outcome	Time to primary failure, defined as A1c ≥ 7% (53 mmol/mol), confirmed	n/a
Years of Study Conduct	2013-2021	2011-2014

Baseline Characteristics of Randomized Cohort		
Demographic		
N		
Age ± SD (years)		
Sex (% male)		
Race/Ethnicity (%) Caucasian African Ancestry Hispanic Asian American Indian		
Duration of diabetes (yr), mean ± SD		
Weight ± SD (kg)		
BMI ± SD (kg/m^2)		
Systolic BP (mmHg)		
Diastolic BP (mmHg)		
Current Smoking (%)		
History of CVD (%)		
Education, years (%) <13 13-16 ≥ 17		

< High school HS graduate Some college ≥ College degree		

Biochemical		
Glycemia		
Fasting Plasma Glucose mg/dL mmol/L		
HbA1c % mmol/mol		
Fasting Insulin pmol/L mU/L		
Lipids		
Total Cholesterol mmol/L mg/dL		
LDL cholesterol mmol/L mg/dL		
HDL cholesterol mmol/L mg/dL		
Triglycerides mmol/L mg/dL		

Table abbreviations: GRADE: Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study; T2DM: type 2 diabetes; FPG: fasting plasma glucose; CVD: cardiovascular disease; LVH: left ventricular hypertrophy; MI: myocardial infarction; CHF: congestive heart failure.

† Non-Hispanic

Table (modified) from: Wexler DJ, Krause-Steinrauf H, Crandall JP, Florez HJ, Hox SH, Kuhn A, Sood A, Underkofler C, Aroda VR, the GRADE Research Group. Baseline Characteristics of Randomized Participants in the Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study (GRADE). Diabetes Care, 2019;11(43):2098-2107.

6.5.6 Table S3: Per-protocol treatment group differences of microvascular outcomes

Same as Table 1 but in per-protocol population.

6.5.7 Table S4: Per-protocol treatment group differences of cardiovascular outcomes

Same as Table 2 but in per-protocol population.

6.6 Figures

6.6.1 List of Figures

Figure	Number	Description
Kaplan-Meier	1	Cumulative incidence of hypertension and hyperlipidemia
Kaplan-Meier	2	Cumulative incidence of microvascular outcomes (ITT)
Kaplan-Meier	3	Cumulative incidence of CVD/mortality outcomes (ITT)
Consort Diagram	S1	Numbers screened and randomized in GRADE by treatment group
Kaplan-Meier	S2	Same as figure 2 but in per-protocol population
Kaplan-Meier	S3	Same as figure 3 but in per-protocol population

6.6.2 Figure 1: Cumulative incidence of hypertension and hyperlipidemia by treatment group

Figure 1 is a two-panel plot showing the cumulative incidence curves for hypertension (left panel) and hyperlipidemia (right-panel) by treatment group. The p-value for the log-rank test of any difference between treatment groups is given at the top of each plot.

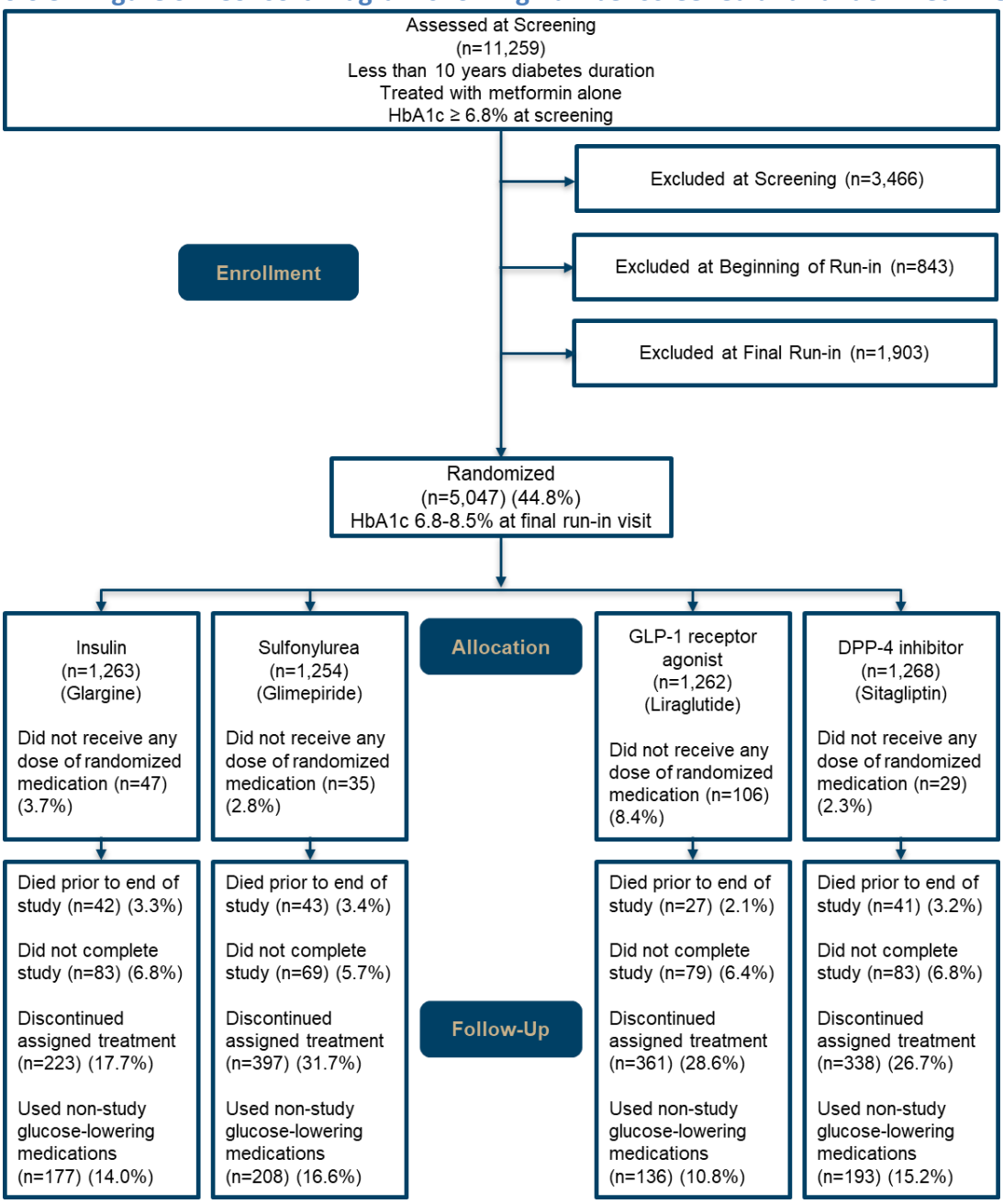
6.6.3 Figure 2: Cumulative incidence of microvascular outcomes (ITT)

Figure 2 is a four-panel plot showing the cumulative incidence curves for moderately increased albuminuria, severely increased albuminuria, impaired eGFR < 60, and neuropathy by treatment group. The p-value for the log-rank test of any difference between treatment groups is given at the top of each plot.

6.6.4 Figure 3: Cumulative incidence of cvd/mortality outcomes (ITT)

Figure 3 is a five-panel plot showing the cumulative incidence curves for any CVD, MACE, hospitalized heart failure, CV death , and all-cause death by treatment group. The p-value for the log-rank test of any difference between treatment groups is given at the top of each plot.

6.6.5 Figure S1: Consort Diagram showing number screened and randomized in GRADE



6.6.6 Figure S2: Cumulative incidence of microvascular outcomes (Per-protocol)

Same as figure 1 but in per-protocol population

6.6.7 Figure S3: Cumulative incidence of cvd/mortality outcomes (Per-protocol)

Same as figure 2 but in per-protocol population

7 STATISTICAL CONSIDERATIONS

7.1 Statistical principles and issues

7.1.1 Significance level of tests

A significance level of α=0.05 will be used for all statistical tests, unless otherwise specified. Comparisons among the treatment groups will be adjusted for the number of tests conducted, 6 for pairwise comparisons and 4 for each group versus the average of the others. Unless stated otherwise, the adjusted p-values are obtained from application of the closed testing principle (7). In cases where the closed testing adjustment cannot be readily applied, then the Holm adjustment will be employed. Otherwise, p-values will be designated as "nominal" or "simple" p-values.

7.1.2 Intention-to-treat analyses

Unless otherwise specified, all available data for all randomized participants (i.e., the full analysis set) will be included in analyses, and data will be analyzed according to the randomly assigned treatment group, regardless of adherence to assigned treatment and/or compliance with the study protocol, according to intention-to-treat principles.

7.1.3 Checking the proportional hazards assumption for the Cox proportional hazards model

For analyses based on the Cox proportional hazards model, the assumption of proportional hazards will be tested using the test of Lin (8). If the test of proportional hazards is significant (i.e., hazards are assessed to be non-proportional), then the coefficients from the Cox model will be interpreted (approximately) as average log hazard ratios, inferences (standard errors, confidence intervals, and p-values) will be based on the robust information sandwich covariance estimates (3), and the robust model score test will be used to test for treatment group differences (2).

7.1.4 Adjustments for multiple pairwise comparisons among the treatment groups

Since there are 4 treatment groups, there are 6 possible pairwise comparisons among the treatment groups. A closed testing approach will be used to account for multiple pairwise comparisons among the treatment groups (7). First, an omnibus T^2-like test will be conducted to test for any differences among the 4 treatment groups. If that test is significant at the specified significance level α, then each of the 3-group sub-hypotheses (i.e., test for differences among 3 of the treatment groups) will be tested at significance level α. Each of the pairwise comparisons can be tested at significance level α if all of the relevant higher-order hypotheses (i.e., 4-group and relevant 3-group hypotheses) are significant at significance level α. See the table below for an outline of the null hypotheses in the testing hierarchy that must be significant to allow for testing of each pairwise comparison. The hypothesis testing tree is given in the table below. Each column gives the series of null hypotheses that are tested in order to establish whether a given order 1 (i.e. 2-group) comparison is significant at level α. For example, the column with header 1 vs 2 gives the series of null hypotheses that are tested (all at level α) in order to establish whether treatment group 1 is significantly different from treatment group 2 at level α. Likewise for the 5 other pairwise group comparisons.

Pairwise Comparison	1 vs 2	1 vs 3	1 vs 4	2 vs 3	2 vs 4	3 vs 4
Order 3(4- group comp)[1]	H_0, 1234	H_0, 1234	H_0, 1234	H_0, 1234	H_0, 1234	H_0, 1234
Order 2(3- group comp)[2]	H_0, 12,13	H_0, 13,12	H_0, 14,12	H_0, 23,12	H_0, 24,12	H_0, 34,12

	$H_0, 12,14$	$H_0, 13,14$	$H_0, 14,13$	$H_0, 23,13$	$H_0, 24,13$	$H_0, 34,13$
	$H_0, 12,23$	$H_0, 13,23$	$H_0, 14,23$	$H_0, 23,14$	$H_0, 24,14$	$H_0, 34,14$
	$H_0, 12,24$	$H_0, 13,24$	$H_0, 14,24$	$H_0, 23,24$	$H_0, 24,23$	$H_0, 34,23$
	$H_0, 12,34$	$H_0, 13,34$	$H_0, 14,34$	$H_0, 23,34$	$H_0, 24,34$	$H_0, 34,24$
Order 1(2- group comp)[3]	$H_0, 12$	$H_0, 13$	$H_0, 14$	$H_0, 23$	$H_0, 24$	$H_0, 34$

[1] $H_0, 1234$ is the null hypothesis that $\mu_1 = \mu_2 = \mu_3 = \mu_4$. This is a 3 df test.

[2] H_0, ab, cd is the null hypothesis that $\mu_a = \mu_b$ and $\mu_c = \mu_d$. These are 2 df tests.

[3] H_0, ab is the null hypothesis that $\mu_a = \mu_b$. These are 1 df tests.

where a,b,c,d ∈ {1,2,3,4}

Note that some of the order 2 tests are equivalent (e.g. $H_0, 13,12 \equiv H_0, 13,23$).

7.1.5 Comparing each treatment to all other treatments combined

There is interest in testing whether the effect of each treatment differs from the other 3 treatment groups combined. Let θ_k be the log(hazard ratio) comparing the hazard for treatment group k=1,2,3 to the hazard for reference treatment group k=4. For each treatment group, we would test the null hypothesis that the average of the estimated hazard ratios comparing each of the other treatments to the treatment of interest equals 1. In other words, we would test each of the following 4 null hypotheses (i.e., one hypothesis per treatment group):

$H_{01}: e^{\theta_2-\theta_1} + e^{\theta_3-\theta_1} + e^{-\theta_1} = 3$

$H_{02}: e^{\theta_1-\theta_2} + e^{\theta_3-\theta_2} + e^{-\theta_2} = 3$

$H_{03}: e^{\theta_1-\theta_2} + e^{\theta_2-\theta_3} + e^{-\theta_3} = 3$

$H_{04}: e^{\theta_1} + e^{\theta_2} + e^{\theta_3} = 3$

A closed testing approach will be used to account for multiple comparisons, according to the procedure described in (4). The closed testing hierarchy would start with the 3-df test of the joint hypothesis $\theta_1 = \theta_2 = \theta_3 = 0$. The next stage of the closed testing hierarchy would be to test the intersections of the elementary hypotheses listed above (e.g., $H_{01} \cap H_{02}$). The last stage would be to test the elementary hypotheses listed above. For example, the elementary hypothesis H_{01} would be rejected at significance level α if $H_{01}, H_{01} \cap H_{02}, H_{01} \cap H_{03}, H_{01} \cap H_{04}$, and the joint hypothesis $\theta_1 = \theta_2 = \theta_3 = 0$ are all significant at significance level α.

7.1.6 Adjustments for multiple comparisons for subgroup analyses

One of the objectives of this paper is to assess treatment group differences within baseline subgroups (e.g., tertiles of HbA1c). There are 6 possible pairwise comparisons among the treatment groups within each subgroup. A Holm adjustment approach will also be used to account for multiple comparisons for testing treatment group differences within subgroups. Here, we describe the approach for the case with all 4 treatment groups and 3 subgroups (e.g., tertiles of HbA1c), where θ_{jk} is the measure of treatment difference between treatment k=1,2,3 and the reference treatment k=4 within subgroup j=a,b,c. First, an

overall test of the null hypothesis of homogeneity of treatment effects across all subgroups would be tested:

H_0: $\theta_{a1} = \theta_{a2} = \theta_{a3}$ AND $\theta_{b1} = \theta_{b2} = \theta_{b3}$ AND $\theta_{c1} = \theta_{c2} = \theta_{c3}$

If this test is significant at the specified significance level (α=0.05), then tests of null hypotheses of homogeneity of the 6 pairwise differential treatment effects is tested:

$H_{0,12}: \theta_{a12} = \theta_{b12} = \theta_{c12}$ [Glargine (k=1) v Glimepiride (k=2)]

$H_{0,13}: \theta_{a13} = \theta_{b13} = \theta_{c13}$ [Glargine (k=1) v Liraglutide (k=3)]

$H_{0,14}: \theta_{a14} = \theta_{b14} = \theta_{c14}$ [Glargine (k=1) v Sitagliptin (k=3)]

$H_{0,23}: \theta_{a23} = \theta_{b23} = \theta_{c23}$ [Glimepiride (k=2) v Liraglutide (k=3)]

$H_{0,24}: \theta_{a24} = \theta_{b24} = \theta_{c24}$ [Glimepiride (k=2) v Sitagliptin (k=4)]

$H_{0,34}: \theta_{a34} = \theta_{b34} = \theta_{c34}$ [Liraglutide (k=3) v Sitagliptin (k=4)]

The p-values from these 6 tests of homogeneity are holm adjusted for 6-tests. The final adjusted p-value for each pairwise test of homogeneity is then taken as the maximum of these holm adjusted p-values and the p-value from the overall test of homogeneity described above.

7.1.7 Calculation of confidence intervals adjusted for multiple comparisons

For analyses with multiple comparisons (e.g., pairwise treatment comparisons, comparisons of each treatment group vs. all others combined, subgroup analyses), confidence intervals for effect estimates will be calculated based on a method that controls the family-wise type 1 error for multiple comparisons.

7.1.8 Confirmation process for Microalbuminuria

The microalbuminuria outcome (ACR $\geq 30\ mg \cdot g^{-1}$) requires confirmation of an initial elevation of ACR with a second elevation at a later visit. In GRADE, routine measurements of ACR occured at semi-annual visits (e.g. visits 00, 06, 12, 18 … etc). However, sometimes there were missed measurements of ACR (due to missed visits or ACR not being measured at a visit) which occurred between two instances of ACR ≥ 30. The confirmation process ignores all missed values of ACR between first (trigger) and second (confirmatory) instances of ACR ≥ 30, provided that there are no intervening measurements of ACR < 30. For example, if a participant has elevated ACR at 30mo, missing values for ACR at 36, 42, 48 months, and elevated ACR again at 54mo, this would count as confirmed microalbumniuria at 30mo. However, if there were a measurement of ACR < 30 at 36, 42 or 48 months, the initial microalbuminuria event would not be confirmed.

8 DISCUSSION POINTS FOR WRITING GROUP

8.1 Considerations

Listing of scientific/statistical issues that need to be discussed by the writing group. For example, how to handle cases where a participant did not have a final lab-based outcome measure because they

experienced a clinical event that is directly related to the lab-based outcome (e.g. did not have an IVGTT because was diagnosed with diabetes prior to the IVGTT visit)

8.2 Limitations

Discuss any limitations of the study (e.g. no baseline measure of outcome, potential selection biases in the study sample etc)

9 APPENDIX A: Dataset Request

9.1 Table of Variables

Measure	Variable	units	Assessment Visits	Notes
Treatment	assign		Baseline	
Age at randomization	age	(yrs)	Baseline	< 45 45-59 60+
Gender	gender	MF.	Baseline	M/F
Race	race	Race.	Baseline	White Black Hispanic Asian Am Indian
Weight	weight	(kg)	Baseline quarterly	
BMI	bmi	(kg/m2)	Baseline quarterly	Categories 22 - < 30 30 - < 35 >=35
SBP	sbp	(mmHg)	Baseline quarterly	
DBP	dbp	(mmHg)	Baseline quarterly	
Hypertension	hyper		Baseline quarterly	Diagnosis OR Use of meds OR Blood pressure >= thresholds on two occasions (140/90) during followup and only once at baseline
Any BP meds	anybp	YN.	Baseline quarterly	
ACEi/ARB	aceiarb	YN.	Baseline quarterly	
Other BP meds	otherbp	YN.	Baseline quarterly	
HDL	hdl	mg/dL	Baseline annual	
LDL	ldl	mg/dL	Baseline annual	
Triglycerides	trig	mg/dL	Baseline annual	

Any Lipid meds	anyllm	YN.	Baseline quarterly	
Statin meds	statins	YN.	Baseline quarterly	
Other Lipid meds	othllm	YN.	Baseline quarterly	
HbA1c	hba1c	mg/dL	Baseline quarterly	
Diabetes Duration	diabdur	yrs	Baseline quarterly	Time from diagnosis to visit
UACR	acr	mg/g	Baseline semi-annual	
UACR ≥ 30	microalb	mg/g	Baseline semi-annual	confirmed (section 7.1.8)
UACR ≥ 300	macroalb	mg/g	Baseline semi-annual	Not confirmed
eGFR	egfr	mL/min/ 1.73m2	Baseline Annual	
eGFR < 60	egfr	mL/min/ 1.73m2	Baseline Annual	Not confirmed
MACE	mace	YN.	Event Time	adjudicated
HF	hf	YN.	Event Time	adjudicated
DPN	DPN	YN.	Baseline Annual	Defined in section 6.2

YN. = Yes or No format (0=No; 1=Yes) MF. = Male or Female format (0=Female; 1=Male)

REFERENCES

1. Herman WH, Pop-Busui R, Braffett BH, et al. Use of the michigan neuropathy screening instrument as a measure of distal symmetrical peripheral neuropathy in type 1 diabetes: Results from the diabetes control and complications trial/epidemiology of diabetes interventions and complications. *Diabetic medicine*. 2012;29(7):937–944.

2. Lachin JM. Biostatistical methods: The assessment of relative risks. 2nd ed. New York: John Wiley & Sons; 2011.

3. Lin DY, Wei LJ. The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association*. 1989;84(408):1074–1078.

4. Lachin JM, Bebu I. Closed testing of each group versus the others combined in a multiple group analysis. *Clinical Trials*. 2020;17(1):77–86.

5. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*. 1986;51(6):1173–1182.

6. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*. 1984;71(3).

7. Lachin JM, Bebu I, Larsen MD, et al. Closed testing using surrogate hypotheses with restricted alternatives. *PLoS ONE*. 2019;(7):1–18.

8. Lin DY. Goodness-of-fit analysis for the cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association*. 1991;86(415):725–728.

Deviations from the Amended Final Statistical Analysis Plan as of April 8, 2022

Treatment group differences in micro/macrovascular outcomes among four initial treatments added to metformin in early type 2 diabetes (OP2)

The original SAP for this paper was signed on May 12, 2021 immediately after the final patient visit in April of 2021. This document summarizes deviations to the original SAP and to the final SAP dated April 8, 2022.

There are 12 deviations are listed below. The first 10 deviations were made to the original SAP of May 12, 2021, the 11th to the final SAP for the original journal submission dated January 19, 2022. The 12th is a new table S2 that was not included in the January 19, 2022 SAP; the final SAP does include this table. Each describes the original text and its source, followed by a description on the deviation and the justification for the change.

List of Deviations:

1. **Original SAP (May 12, 2021)**: The original SAP describes six outcomes for the paper (section 6.2) including two CVD outcomes (MACE, hospitalized heart failure), and four microvascular outcomes (eGFR < 60, confirmed ACR ≥ 30 mg/g, ACR ≥ 300 mg/g, and distal symmetric polyneuropathy (DSPN).

   Deviation: The paper describes results for an additional three outcomes: a composite CVD outcome (any CVD), cardiovascular death alone, and total mortality.

   Justification:

   - Excluding any CVD was an oversight when writing the SAP. The GRADE study protocol describes the components of cardiovascular disease (Section 8.1) included in the definition of the any CVD composite outcome presented in the results (MACE, hospitalized heart failure, unstable angina requiring hospitalization, or any revascularization procedure). After realizing that the any CVD composite outcome was not part of the SAP, John Lachin sent an email (dated July 6, 2021) to the GRADE executive committee pointing out the omission, and documenting early drafts of the SAP that included a composite CVD outcome. The decision to amend the SAP was made.
   - The CV mortality outcome is a component of MACE, and it was decided that it should be reported individually.
   - The total mortality outcome was added to the ADA presentation per the suggestion of John Buse (May 23, 2021).

2. **Original SAP (May 12, 2021)**: Dialysis/Transplant/Renal death were not included as part of the microvascular outcomes (confirmed ACR ≥ 30 mg/g, and ACR ≥ 300 mg/g)

   Deviation: Dialysis/Transplant/Renal Death are now included as an additional condition for defining the albuminuria outcomes

   Explanation:
   - An oversight when creating the outcome variables
   - There are very few occurrences of dialysis/transplant/ESRD death (N=10) and results did not change appreciably

3. **Original SAP (May 12, 2021)**: Figure 1 was changed from a 6 panel plot with KM curves for MACE, HF, DSPN, eGFR < 60, confirmed microalbuminuria, and macroalbuminuria outcomes to a 2-panel plot with KM incidence curves for hypertension and dyslipidemia.

   Deviation: The Kaplan-Meier curves for hypertension and dyslipidemia were put into a separate figure (Figure 1 in the revised SAP). DSPN was corrected to DPN for Diabetic Peripheral Neuropathy.

   Justification: Hypertension and dyslipidemia are important risk factors for macro-vascular and micro-vascular outcomes (principally macro-vascular, where they are strongly related to the outcomes). It was decided that an assessment of treatment group differences in their incidence would be helpful when interpreting results. DSPN was incorrect and corrected to DPN to be consistent with ADA guidelines.

4. **Original SAP (May 12, 2021)**: The original SAP for this paper included calculations of the RMST within each group along with pairwise tests of significance among groups.

   Deviation: Owing to concerns of redundant tests of pairwise group differences, and negligible differences in RMST between groups for these outcomes, it was decided not to present RMST values in tables 1 and 2.

   Justification: RMST difference is not informative when there is no difference between groups. To avoid having two tests of treatment group differences (hazard ratios and RMST differences) that may conflict, it was decided instead to delete RMST calculations.

5. **Original SAP (May 12, 2021)**: Mediation analyses were included as part of the original SAP (Table 4)

   Deviation: Mediation analyses were excluded from this manuscript (Table 4 removed)

   Justification: Based on discussion with the writing group, it was decided (prior to conducting any mediation analyses) that mediation analyses would be too complex to add to this paper, given the amount of content already included in this manuscript, and that it would be better to address mediation analyses in another manuscript.

6. **Original SAP (May 12, 2021)**: For subgroup analyses, adjustment for multiple comparisons was described using a closed testing procedure (see description for table 5 on page 9).

   Deviation: The Holm procedure was used to protect type 1 error due to multiple testing instead of closed testing described in the original SAP.

   Justification: The closed testing procedure for this setting is not available in existing statistical software, so a more conservative holm adjustment (readily available in software) was implemented.

7. **Original SAP (May 12, 2021)**: Figure 2 in the original SAP was a series of heterogeneity plots by subgroup factors.

   Deviation: This plot is eliminated and replaced by figures 2 and 3 in the revised SAP, KM incidence curves of microvascular and DPN (Figure 2) and cardiovascular/death (Figure 3) outcomes

   Justification: Heterogeneity results showed few subgroup differences so the figure was used to instead split out the cardiovascular and microvascular KM plots since several additional outcomes were added (any CVD, CV deaths, all deaths). Notable subgroup differences are described in the manuscript text.

8. **Original SAP (May 12, 2021)**: Per-protocol analyses were not included as part of the original SAP

   Deviation: Per-protocol sensitivity analyses were added in the final SAP (March 16, 2022). See tables S2, S3 and figures S2, S3.

   Justification: The per-protocol sensitivity analyses are important to assess whether any observed ITT treatment differences on outcomes might be explained by differences in adherence to randomized treatments.

9. **Original SAP (May 12, 2021)**: For subgroup analyses, race and ethnicity were combined in a single variable (non-Hispanic White, non-Hispanic Black, Hispanic White, Other).

   Deviation: race (White, Black, and Other/Multiple) and ethnicity (Hispanic/Latino vs. not Hispanic/Latino) were considered as separate subgroup variables.

   Justification: Based on discussion with the writing group, it was decided that race and ethnicity are separate constructs, and so should be analyzed separately.

10. **Original SAP (May 12, 2021)**: Objective 3 describes a comparison of longitudinal risk factors for micro/macrovascular outcomes at landmark time points of 1-year and 5-years post randomization (Section 5.2, objective 3). This was correct in the 3rd aim of the SAP (Section 6.4.3).

    Deviation: In section 5.2 of the SAP (04/08/2022) for the current re-submission, objective 3 is corrected to indicate landmark analyses will be conducted at years 1 and 4.

Justification: **At 5 years of follow-up 59% of the cohort of 5047 remained under follow-up, the substantial fraction missing being largely a function of administrative curtailment of follow-up owing to staggered entry into the trial. At 4 years of follow-up, 86% of the cohort remained under follow-up that would provide more accurate estimates of group differences in outcomes than would be the case at 5 years. Accordingly the writing committee decided to change the analysis time point from 5 years to 4 years of follow-up.**

11. **SAP version (January 19, 2022):** The definition of prevalent and incident hypertension used SBP/DBP thresholds of 130/80 mmHg (see section 6.4.0), with the full definition being the following:

    1. Prevalent Hypertension at baseline:
        - systolic BP ≥130 mmHg OR diastolic BP ≥80 mmHg OR
        - self-reported use of blood pressure-lowering medications for control of blood pressure OR
        - self reported diagnosis of hypertension between at screening or baseline visit
    2. Incident hypertension
        - No prevalent hypertension AND
        - (systolic BP ≥130 mmHg OR diastolic BP ≥80 mmHg) at visit confirmed at a subsequent visit OR
        - self-reported use of blood pressure lowering medications for control of blood pressure since last visit OR
        - self reported diagnosis of hypertension since last visit

    Deviation: The final SAP (April 8, 2022) that accompanies this re-submission uses ADA thresholds of 140/90 in the definition of prevalent and incident hypertension (see section 6.4.0), with the full definition as follows:

    1. Prevalent Hypertension at baseline:
        - systolic BP ≥140 mmHg OR diastolic BP ≥90 mmHg OR
        - self-reported use of blood pressure-lowering medications for control of blood pressure OR
        - self reported diagnosis of hypertension at the screening or baseline visit.
    2. Incident hypertension
        - No prevalent hypertension AND
        - (systolic BP ≥140 mmHg OR diastolic BP ≥90 mmHg) at visit confirmed at a subsequent visit OR
        - self-reported use of blood pressure lowering medications for control of blood pressure since last visit OR
        - self reported diagnosis of hypertension since last visit

    Justification: During the entire planning and conduct of GRADE, we have consistently *stated* in the protocol that the outcome of hypertension was defined as:

- Incidence and prevalence of hypertension defined as blood pressure of ≥140 mmHg systolic or ≥90 mmHg diastolic OR the use of blood pressure-lowering medications for control of blood pressure
- Incidence of emergent hypertension among those who had levels <140/90 and were free of blood pressure-lowering medication use at baseline.

Unfortunately, the SAP (01/19/22), changed the definition, which was an error, to systolic ≥130 or diastolic ≥80 mmHg. The analyses use the BP threshold of ≥140/90 as noted in the protocol, which needs to be confirmed at a subsequent visit in accord with the ADA definition of hypertension. We realized the discrepancy in definitions and revised the SAP prior to performing any analyses with the new (corrected) definitions of hypertension.

12. **Final SAP for initial submission (January 19, 2022):** The prior SAP did not include this table S2, Table S2: Comparison of characteristics in the GRADE and NHANES populations. This table was requested by the reviewers and is included in the final SAP (April 8, 2022).