CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

215866Orig1s000

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDA 215866
Supplement #:	S001 (Seq. No.: 001; SDN: 1)
Drug Name:	MOUNJARO™ (tirzepatide) Injection
Indication(s):	Adjunct to diet and exercise to improve glycemic control in T2DM
Applicant:	Eli Lilly and Company
Date(s):	Stamp Date: 9/15/2021
	Date received by reviewer: 12/8/2021
	PDUFA due date: 5/15/2022
Review Priority:	Standard
Biometrics Division:	DB VII
Statistical Reviewer:	Shanti Gomatam, Ph.D.
Concurring Reviewers:	Statistical Team Lead: Eugenio Andraca-Carrera, Ph.D.
	Statistical Division Director: Mark Levenson, Ph.D.
Medical Division:	OCHEN/DDLO
Clinical Team:	Clinical Reviewer/Medical Officer: Frank Pucino, M.D.
	Clinical Team Lead: Michael Nguyen, M.D.
	Medical Division Director: John Sharretts, M.D.
Project Manager:	Lindsey Kelly
	Team Leader: Callie Cappel-Lynch

Keywords: MACE, all-cause mortality, pulse rate, meta-analysis, Survival, Cox Proportional Hazards, stratified, linear model

Table of Contents

1 EXECUTIVE SUMMARY 5

1.1 CONCLUSIONS AND RECOMMENDATIONS 5
1.2 STATISTICAL ISSUES AND FINDINGS 6

2 INTRODUCTION 8

2.1 OVERVIEW 9
2.2 DATA SOURCES 11

3 STATISTICAL EVALUATION 11

3.1 DATA AND ANALYSIS QUALITY 11
3.2 EVALUATION OF EFFICACY 11
3.3 EVALUATION OF CV SAFETY 11
3.3.1 CV Meta-Analysis Trial Designs and Endpoints 11
3.3.1.1 Trial Designs 12
3.3.1.2 Treatment Regimens in the Trials 16
3.3.1.3 Endpoints 16
3.3.1.4 Sample Size Calculation 16
3.3.1.5 Clinical Events Adjudication 17
3.3.2 Statistical Methodology 17
3.3.2.1 Analysis Populations 17
3.3.2.2 Analysis Methods for Safety Endpoints 17
3.3.2.3 Interim Analyses and Multiple Testing Plan 18
3.3.3 Patient Disposition, Demographics and Baseline Characteristics 18
3.3.3.1 Patient Follow-up and Disposition 19
3.3.3.2 Demographics and Baseline Characteristics 21
3.3.4 Results and Conclusions 23
3.3.4.1 Primary Safety Endpoint Analyses 23
3.3.4.2 Supporting analyses for Primary MACE endpoint 27
3.3.4.3 All-cause mortality Analyses 31
3.3.4.4 Exploratory Analyses for Pulse Rate Changes 33

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS 36

4.1 GENDER, RACE, AGE, AND GEOGRAPHIC REGION SUBGROUPS FOR MACE-4 ENDPOINT . 36
4.2 CLINICAL SUBGROUPS FOR MACE-4 ENDPOINT 37
4.3 GENDER, RACE, AGE, AND GEOGRAPHIC REGION SUBGROUPS FOR ALL-CAUSE MORTALITY 39
4.4 CLINICAL SUBGROUPS FOR ALL-CAUSE MORTALITY 41

5 SUMMARY AND CONCLUSIONS 43

5.1 COLLECTIVE EVIDENCE 43
5.2 STATISTICAL ISSUES 43
5.3 CONCLUSIONS AND RECOMMENDATIONS 44

Reference ID: 4972437

LIST OF TABLES

Table 1: Analysis of pre-specified primary MACE-4 endpoint in CV meta-analysis by tirzepatide dose (mITT; IOMA)* 7
Table 2: Pre-specified Analysis of all-cause mortality Endpoint (mITT, EOMA)* 7
Table 3: Overview of T2DM studies contributing to CV meta-analysis 10
Table 4: Description of tirzepatide studies included in the CV meta-analysis 13
Table 5. Subject follow-up in years by stratum and treatment arm (ITT) 20
Table 6: Discontinuation reasons by Study (ITT) 21
Table 7: Baseline Demographic Characteristics (ITT) 22
Table 8: Baseline Clinical Characteristics (ITT) 23
Table 9: Patients with MACE-4 events by Trial (mITT, IOMA)* 24
Table 10: Analysis of primary MACE-4 endpoint in CV meta-analysis by tirzepatide dose (mITT; IOMA)* 25
Table 11: Description of components of primary MACE-4 endpoint in CV meta-analysis (mITT; IOMA)* 27
Table 12: Patients with MACE-4 events by Trial (mITT, EOMA)* 28
Table 13: Analysis of primary MACE-4 endpoint in CV meta-analysis by tirzepatide dose (mITT; EOMA)* 29
Table 14: Description of MACE-4 endpoint and components (mITT, EOMA)* 31
Table 15: Pre-specified Analysis of all-cause mortality Endpoint (mITT, EOMA)* 32
Table 16: Number of patients by study and dose in meta-analysis Phase 3 studies 33
Table 17: Mean and standard deviation of pulse at baseline and Week 24 by trial and Japan/non-Japan subgroups in meta-analysis Phase 3 studies* 34
Table 20: Summary of Results for model of Week 24 change in pulse rate from baseline for Phase 3 studies included in the meta-analysis including main effect for Japan 34
Table 19: Summary of Results for models of Week 24 change in pulse rate from baseline for Phase 3 studies included in the meta-analysis 35
Table 21: Demographic Subgroup analyses for MACE-4 endpoint (mITT; EOS) 36
Table 22: Clinical subgroup analyses for MACE-4 endpoint (mITT; EOS) 38
Table 23: Demographic Subgroup analyses for all-cause mortality endpoint (mITT; EOS) 39
Table 24: Clinical subgroup analyses for ACM endpoint (mITT; EOS) 41

Reference ID: 4972437

LIST OF FIGURES

Figure 1: Follow-up by treatment arm for Trial GPGM (ITT) 19
Figure 2: Follow-up by treatment arm for trials other than GPGM (ITT)..... 20
Figure 3: Kaplan-Meier-based cumulative incidence plots for MACE-4 by treatment arm (mITT, IOMA)*..... 26
Figure 4: Kaplan-Meier-based cumulative incidence plots for MACE-4 by treatment arm and dose (mITT, IOMA)*..... 26
Figure 5: Kaplan-Meier-based cumulative incidence plots for MACE-4 by treatment arm (mITT, EOMA)* 30
Figure 6: Kaplan-Meier-based cumulative incidence plots for MACE-4 by treatment arm and dose (mITT, EOMA)* 30
Figure 7: Kaplan-Meier-based cumulative incidence plots for all-cause mortality by pooled treatment arms (mITT, EOMA)* 32
Figure 8: Kaplan-Meier-based cumulative incidence plots for all-cause mortality by treatment arms and tirzepatide dose arms (mITT, EOMA)* 33
Figure 9: Forest plot for demographic Subgroups for MACE-4 endpoint (mITT; EOS)..... 37
Figure 10: Forest plot for clinical subgroups for MACE-4 endpoint (mITT; EOS)..... 39
Figure 11: Forest plot for demographic subgroups for all-cause mortality (mITT; EOS)..... 41
Figure 12: Forest plot for clinical subgroups for all-cause mortality (mITT; EOS)..... 42

Reference ID: 4972437

1 EXECUTIVE SUMMARY

MOUNJARO™ (LY3298176; generic name: tirzepatide) is an incretin mimetic that binds to both the glucose-dependent insulinotropic polypeptide (GIP) and to glucagon-like peptide-1 (GLP-1) receptor agonist with proposed indication once-weekly (QW) as an adjunct to diet and exercise to improve glycemic control in adults with type 2 diabetes mellitus.
This is a review of the pre-specified cardiovascular (CV) meta-analysis conducted by the applicant. In addition, exploratory analyses were conducted to assess clinical concerns of differences in pulse rate changes. Data from 7 clinical studies were analyzed in this meta-analysis to investigate the efficacy and safety of tirzepatide 1 mg, 5 mg, 10 mg and 15 mg once-weekly.

1.1 Conclusions and Recommendations

The cardiovascular safety of tirzepatide was evaluated based on results of the CV meta-analysis that included 7 trials: 1 Phase 2 trial and 6 Phase 3 trials. Three of the 7 trials were open-label trials whereas the rest were double-blind. The comparator arms of these trials included placebo and active comparators. A total of 7215 patients was randomized and took at least one dose of study treatment. These subjects contributed a total of 7781.8 patient-years of on-study follow-up time – 5064.45 on the tirzepatide arms and 2717.35 on the pooled comparator arms.

The primary safety endpoint was the 4-component major adverse cardiovascular event (MACE-4) defined as a composite of the following adjudicated events: cardiovascular death, myocardial infarction, stroke, and hospitalization for unstable angina. The applicant conducted a pre-specified interim analysis of MACE-4 in the set of 7 trials after 116 MACE-4 events were observed. The estimated hazard ratio was 0.81 with an associated 97.85% CI of (0.52, 1.26) – the 97.85% level was the confidence level obtained from the pre-specified Hwang, Shih, DeCani spending function. The upper bound of this confidence interval is less than 1.8 and meets the risk margin pre-specified in the meta-analysis Statistical Analysis Plan.

A hazard ratio of 0.80 with an associated 95%[1] CI of (0.51, 1.25) was estimated for the all-cause mortality endpoint for the comparison of pooled tirzepatide arms versus pooled comparator arm based on final on-study follow-up. This confidence interval includes the null value of 1 and does not raise any concerns of excess mortality in the tirzepatide arm.

Exploratory analyses were conducted to assess pulse rate changes, in the subset of patients randomized in Japan over all trials, at the request of the clinical review team. Subjects in Japan observed a higher average change from baseline in pulse rate than subjects in other countries. Exploratory analyses of Week 24 change from baseline in pulse rate indicate that Dose, baseline Pulse, and their interaction are significant predictors of Week 24 change in pulse rate. However, there are also differences in Japan and non-Japan subsets – for the Japan subgroup baseline Pulse

[1] Note that no interim analysis was planned for the all-cause mortality endpoint, hence no adjustment to the confidence level was required. This endpoint was analyzed using the final data at the 95% confidence level.

Reference ID: 4972437

is the only significant factor among these three. Pre-specification of hypotheses for future studies is recommended for confirmatory conclusions.

Based on this evidence, we consider that the CV meta-analysis was generally successful in demonstrating cardiovascular safety of tirzepatide when compared to the standard of care. The meta-analysis included open label studies and different active comparators, hence these results should be interpreted and generalized with caution.

1.2 Statistical Issues and Findings

The following issues should be considered when interpreting the results of the meta-analysis provided to assess CV risk of tirzepatide:

- 3 of the 7 randomized trials in the meta-analysis were open-label trials; knowledge of treatment assignment could cause bias in results due to differences in patient management and other factors;
- Different comparators – active and placebo – were used in the different studies in the meta-analysis. Study GPGM which contributed most of the MACE-4 events had insulin glargine as the active comparator and showed a numerically beneficial effect for MACE-4 in the tirzepatide arm; Study GPGL which had semaglutide as the active comparator, showed a numerically beneficial effect in the semaglutide arm. Any potential generalization of results from this meta-analysis would have to take this into account.

Table 1 contains results from the pre-specified analysis for the primary MACE-4 endpoint that includes all adjudicated MACE-4 events in the mITT population, i.e., all randomized subjects who took at least one dose of study treatment. Per the Statistical Analysis Plan, events that occurred up to 30 days after the treatment period, and were positively adjudicated by the CEC, were to be included in the analysis. A hazard ratio of 0.81 was estimated for the MACE-4 endpoint. The upper bound of the pre-specified 97.85% confidence interval at the interim analysis for the hazard ratio for the pooled tirzepatide doses versus all comparators was 0.81; this upper bound was lower than the1.8 risk margin pre-specified for the non-inferiority hypothesis on the MACE-4 endpoint. Thus, it can be concluded that the analysis for MACE successfully ruled out the 1.8 risk margin.
Analyses of the end-of-meta-analysis (complete data) and of the components of MACE-4 were supportive of this conclusion, as were the subgroup analyses.

Reference ID: 4972437

Table 1: Analysis of pre-specified primary MACE-4 endpoint in CV meta-analysis by tirzepatide dose (mITT; IOMA)*

	Tirzepatide					Pooled Comparator
	15 mg N= 1461	**10 mg N=1448**	**5 mg N= 1449**	**1 mg N=52**	**All N= 4410**	**N=2169**
MACE-4 patient years of follow-up	1457.19	1435.64	1428.04	28.59	4349.45	2344.46
MACE-4 events (rate per 100 PY)	16 (1.10)	18 (1.25)	25 (1.75)	1 (3.50)	60 (1.38)	56 (2.39)
HR (97.85% CI)§	0.64 (0.33, 1.24)	0.74 (0.40, 1.39)	1.04 (0.59, 1.82)	5.01 (0.46, 54.75)	0.81 (0.52, 1.26)	-

* **mITT**: All randomized subjects who took at least one dose of study drug; **IOMA**: Interim look of the meta-analysis
‡ Hazard ratios for pooled tirzepatide doses computed using a Cox proportional hazards model with fixed effects for treatment and stratified by study-level CV Risk.
§ The confidence level of 97.85 was based on the 116 events observed at interim and the pre-specified Hwang, Shih and DeCani spending function with Gamma=-6.6.

Source: Created by FDA statistical reviewer using adtte provided by applicant in February 2022.

Analyses for the all-cause mortality endpoint on the mITT population were pre-specified but not covered under the multiple testing plan. This analysis, presented in Table 2, resulted in a hazard ratio estimate of 0.80 with a 95% confidence interval of (0.51, 1.25). This interval covers the null value of 1 and does not raise any concerns about excess mortality in the tirzepatide arms.

Table 2: Pre-specified Analysis of all-cause mortality Endpoint (mITT, EOMA)*

	Tirzepatide					Pooled Comparator
	15 mg N=1621	**10 mg N=1606**	**5 mg N= 1608**	**1 mg N=52**	**All N= 4887**	**N=2328**
ACM patient years of follow-up	1708.29	1685.82	1676.61	28.67	5099.16	2756.29
ACM events (IR per 100 PY)	13 (0.76)	8 (0.47)	20 (1.19)	0 (0.00)	41 (0.80)	39 (1.41)
HR (95% CI)	0.75 (0.40, 1.41)	0.47 (0.22, 1.01)	1.20 (0.69, 2.08)	0 (0, Inf)	0.80 (0.51, 1.25)	-

* **mITT**: All randomized subjects who took at least one dose of study drug; **EOMA**: end-of-meta-analysis.
‡ Hazard ratios for pooled tirzepatide doses computed using a Cox proportional hazards model with fixed effects for treatment and stratified by study-level CV risk (GPGM versus others)

Source: Created by FDA statistical reviewer.

Reference ID: 4972437

2 INTRODUCTION

MOUNJARO™ (LY3298176; generic name: tirzepatide) is an incretin mimetic that binds to both the glucose-dependent insulinotropic polypeptide (GIP) and a glucagon-like peptide-1 (GLP-1) receptor agonist. The applicant proposes a once-weekly (QW) indication as an adjunct to diet and exercise to improve glycemic control in adults with type 2 diabetes mellitus.
This is a review of the cardiovascular (CV) meta-analysis conducted by the applicant. Data from 7 clinical studies were included in this CV meta-analysis to investigate the efficacy and safety of tirzepatide once-weekly.

The applicant's meta-analysis was planned in accordance with the 2008 FDA guidance for industry[2] – Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes Mellitus (T2DM). The meta-analysis of Phase 2 and pivotal Phase 3 registration trials conducted in T2DM patients was to assess whether the 1.8 CV risk margin specified in the 2008 guidance was met by tirzepatide before regulatory submission for the T2DM indication. A 4-component major adverse CV event (MACE-4) endpoint consisting of: death due to CV cause, myocardial infarction (MI), stroke, and hospitalization for unstable angina (HUA), was used for this assessment of tirzepatide versus a pooled comparator arm that included placebo, dulaglutide 0.75 mg and 1.5 mg doses, semaglutide 1.0 mg, insulin degludec and insulin glargine. The applicant intends to use this meta-analysis to support the initial approval of tirzepatide.

The applicant is also conducting a large Phase 3 CV outcomes trial (CVOT), called SURPASS-CVOT, with an expected enrollment of 12500 patients with a history of CV disease and an average follow-up of 4 years. In this CVOT, patients will be randomized using a 1:1 ratio to a QW dosing regimen of tirzepatide dose up to 15 mg or to dulaglutide 1.5 mg. The primary objective of the trial is to demonstrate that the tirzepatide dosing regimen is noninferior to the dulaglutide 1.5 mg dosing regimen relative to the composite 3-component MACE (MACE-3) endpoint consisting of: CV death, MI, and stroke.

Below are the Indications and dosage information from the proposed label based on the package insert submitted by the sponsor with the September 15, 2021, submission.

Indication and Usage:
MOUNJARO™ is a dual glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide-1 (GLP-1) receptor agonist indicated as an adjunct to diet and exercise to improve glycemic control in adults with type 2 diabetes mellitus. (1)
Limitations of Use:

- Has not been studied in patients with a history of pancreatitis and should be used with caution in these patients (1, 5.2)
- (b) (4) type 1 diabetes mellitus (1)

[2] This is no longer a requirement for antidiabetic therapies to treat Type 2 Diabetes Mellitus. The draft guidance for industry "Type 2 Diabetes Mellitus: Evaluating the Safety of New Drugs for Improving Glycemic Control", March 2020, provides recommendations on the size and nature of the safety databases needed to support drugs for chronic use to improve glycemic control in patients with type 2 diabetes.

Reference ID: 4972437

(b) (4)

Reviewer Comment:
Of the 6 doses mentioned in Dosage Forms and Strengths above, the 2.5 mg, 7.5 mg and 12.5 mg doses do not appear to have been assigned to any patients in the studies included in the CV meta-analysis.

2.1 Overview

Dr. Frank Pucino is the clinical reviewer for this submission and the statistical review of efficacy aspects is being conducted by Dr. Wenda Tu of DBII. IND 128801, under which the studies for the meta-analyses were conducted, was reviewed by Dr. Bo Li of DBVII and Drs. Jennifer Clark and Wenda Tu of DB II. The SAP for the CV meta-analysis was reviewed by Dr. Bo Li. Reviews by DBVII included those of the meta-analysis SAP (submitted 9/28/20), the EOP2 meeting package (submitted 7/17/18) and SAP and DMC Charter for trial GPGM (submitted 12/21/18).

A meta-analysis of Phase 2 and Phase 3 trials was used to assess CV risk for tirzepatide. Trials included in the CV risk assessment were I8F-MC-GPGB, I8F-MC-GPGK, I8F-MC-GPGL, I8F-MC-GPGH, I8F-MC-GPGM and I8F-MC-GPGO.

Only the CV risk assessment carried out using the meta-analysis is being reviewed here. Trials to be included in the meta-analysis per the applicant's Statistical Analysis Plan (SAP) are described in Table 3. As pre-specified in the Meta-analysis Statistical Analysis Plan, Phase 1 clinical trials conducted in T2DM patients and the Phase 2 study I8F-MC-GPGF[3] were excluded from the meta-analysis because they were trials with shorter duration and did not result in any MACE-4 events. A Phase 3 trial I8F-JE-GPGP in Japanese T2DM patients was excluded as it did not include a control arm.

[3] Trial GPGF had a 2-week screening/lead-in period, a 12-week treatment period and a 4-week safety follow-up period.

Reference ID: 4972437

Table 3: Overview of T2DM studies contributing to CV meta-analysis

Study ID (Background Treatment)	Description	Design	Treatment Duration[a]	Treatments
Phase 2 Glycemic Control Studies				
I8F-MC-GPGB (with or without metformin)	H2H vs placebo, dulaglutide	Randomized, double-blind, placebo-controlled and active-controlled	26 wks	• TZP 1 mg • TZP 5 mg • TZP 10 mg • TZP 15 mg • Placebo • Dulaglutide 1.5 mg
Phase 3 Glycemic Control Studies				
I8F-MC-GPGK (none)	H2H vs Placebo (monotherapy)	Randomized, double-blind, placebo-controlled	40 wks	• TZP 5 mg • TZP 10 mg • TZP 15 mg • Placebo
I8F-MC-GPGL (add on to metformin)	H2H vs semaglutide	Randomized, open-label, active-controlled	40 wks	• TZP 5 mg • TZP 10 mg • TZP 15 mg • Semaglutide 1.0 mg
I8F-MC-GPGH (add on metformin±SGLT2i)	H2H vs insulin degludec	Randomized, open-label, active-controlled	52 wks	• TZP 5 mg • TZP 10 mg • TZP 15 mg • Insulin degludec
I8F-MC-GPGM (add on to 1-3 OAMs)	H2H vs insulin glargine	Randomized, open-label, active-controlled	52 wks min-104 wks max	• TZP 5 mg • TZP 10 mg • TZP 15 mg • Insulin glargine
I8F-MC-GPGI (add on to basal insulin with or without metformin)	H2H vs Placebo	Randomized, double-blind, placebo-controlled	40 wks	• TZP 5 mg • TZP 10 mg • TZP 15 mg • Placebo
Phase 3 Glycemic Control Studies				
I8F-JE-GPGO[b] (none)	H2H vs dulaglutide (monotherapy)	Randomized, double-blind, active-controlled	52 wks	• TZP 5 mg • TZP 10 mg • TZP 15 mg • Dulaglutide 0.75 mg

Abbreviations: CV = cardiovascular; H2H = head to head; ID = identifier; max = maximum; min = minimum; OAM = oral antihyperglycemic medication; T2DM = type 2 diabetes mellitus; TZP = tirzepatide; wks = weeks.

a All patients have a follow-up period of 4 weeks after treatment period.

b Will be included in interim or final meta-analysis if the study is completed at the time of meta-analysis.

Source: Applicant's Cardiovascular Meta-Analysis Plan Version 2, Table 4.1 (page 7/21).

A vast majority of the MACE events for the CV safety analysis were anticipated from Trial 18F-MC-GPGM (Trial GPGM). Whereas the other 6 trials were conducted in patients with expected CV risk associated with T2DM, Study GPGM enrolled patients with increased CV risk (defined as patients with coronary artery disease, peripheral arterial disease, cerebrovascular disease, chronic heart failure, and chronic kidney disease) resulting in an enriched pool of patients in the elderly age group (around 47.5% who were 65 years or older) and those with longer duration of T2DM (mean duration of diabetes 11.8 years). Trial GPGM was also the largest and longest Phase 3 study in the tirzepatide clinical program.

Reference ID: 4972437

We discuss in this review of the CV meta-analysis, the MACE-4 and all-cause mortality related objectives, endpoints and analyses, which are of interest in this statistical review of safety. An exploratory review of pulse rate changes was also conducted at the request of the clinical review team and is summarized in this review.

Henceforth the last 4 letters of the Study IDs will be used in most instances to refer to the trials – for example, Trial I8F-MC-GPGB will be referred to as Trial GPGB.

2.2 Data Sources

The submission and associated data were provided electronically.
The submission can be found at:
\\CDSESUB1\evsprod\NDA215866\0001

ADaM data sets provided by the sponsor for the meta-analysis were used for the statistical analyses conducted in this review. Refer links below for datasets:
\\CDSESUB1\evsprod\NDA215866\0001\m5\datasets

The applicant also provided additional data on February 11, 2022 in response to an information request (IR) for the MACE-4 interim analysis. These data can be found at:

\\CDSESUB1\evsprod\NDA215866\0035\m5\datasets.

3 STATISTICAL EVALUATION

A statistical review of quantitative safety for this original supplement is presented below.
This review focuses on cardiovascular safety assessment based on the applicant's meta-analysis.
At the request of the clinical team an exploratory analysis of pulse rate was also conducted.

3.1 Data and Analysis Quality

The data and reports for this submission were submitted electronically. The electronic data provided in the original submission and in response to the IR were of adequate quality to conduct statistical analyses.

3.2 Evaluation of Efficacy

Refer to the review by Dr. Wenda Tu of OB/DBII for a statistical review of the efficacy aspects of this submission.

3.3 Evaluation of CV Safety

3.3.1 CV Meta-Analysis Trial Designs and Endpoints

The **primary objective** of the CV meta-analysis was to demonstrate that tirzepatide was not associated with an unacceptably high risk for 4-component MACE (MACE-4) – a composite endpoint with death due to CV causes, myocardial infarction (MI) and stroke – in patients with Type 2 Diabetes Mellitus (T2DM). This objective was to be evaluated by comparing the distribution of time to first occurrence of MACE-4 for patients receiving any dose of QW tirzepatide (pooled tirzepatide group) to that in patients administered comparators, placebo or

Reference ID: 4972437

active control (pooled control group). The primary objective would be considered to have been met if the upper bound of the 2-sided 95% confidence interval (CI) for the hazard ratio (HR) for pooled tirzepatide versus pooled control from the meta-analysis was less than 1.8.

Additional objectives are comparison of time to first occurrence between pooled tirzepatide group to pooled control group for the following endpoints:

- All-cause mortality
- 3-component MACE consisting of death due to CV causes, MI and stroke (MACE-3)
- CV death
- The composite endpoint of MACE-3 or hospitalization for heart failure
- All MI
- All stroke
- HUA
- Hospitalization for heart failure

3.3.1.1 Trial Designs

Per the applicant's Statistical Analysis Plan (SAP) the meta-analysis was to include at least one Phase 2 study (Study 18F-MC-GPGB) and 5 multi-regional Phase 3 trials (Studies 18F-MC-GPGH, 18F-MC-GPGI, 18F-MC-GPGK, 18F-MC-GPGL and 18F-MC-GPGM). The final version of the SAP (Version 2) states that only data from Phase 2 and Phase 3 studies with at least 26 weeks planned treatment duration and a placebo and/or active comparator were to be included in the CV safety meta-analysis.

A description of the design of the trials in the pooled meta-analysis of MACE-4 is in Table 4 below. Six trials were included in the interim analysis for the MACE 4 endpoint – Trial GPGO was not included in the interim analysis since database lock for this trial had not occurred when the interim analysis was conducted. However, all 7 trials in Table 4 were included in end-of-meta-analysis/complete data analyses.

Reference ID: 4972437

Table 4: Description of tirzepatide studies included in the CV meta-analysis

Study	Patient Population (Key Inclusion Criteria)	Background Therapy	Study Design	Control/ Comparator	Treatment Duration[a]	No. of Patients Treated with Study Drug by Treatment (Randomization Ratio)
Phase 2 Study						
I8F-MC-GPGB	M or F, 18-75 years of age, inclusive, with T2DM for ≥6 months, on diet and exercise ± metformin for ≥3 months; HbA1c of 7.0%-10.5%, inclusive; BMI of 23-50 kg/m^2, inclusive; eGFR ≥45 mL/min/1.73 m^2.	±Metformin for ≥3 months	Randomized, double- blind, placebo-controlled and active-controlled	Placebo Dulaglutide 1.5 mg	26 weeks	TZP 1 mg QW 52 TZP 5 mg QW 55 TZP 10 mg QW 51 TZP 15 mg QW 53 Placebo 51 Dula 1.5 mg QW 54 Total 316 (1:1:1:1:1:1)
Global Phase 3 Studies						
I8F-MC-GPGK (monotherapy)	M or F, ≥18 years of age with T2DM, naive of diabetes injectable therapies and no OAM use ≥3 months prior to study entry; HbA1c of 7.0%-9.5%, inclusive; stable weight ≥3 months prior to study entry; eGFR ≥30 mL/min/1.73 m^2; BMI ≥23 kg/m^2.	No OAM use ≥3 months prior to study entry	Randomized, double-blind, placebo-controlled	Placebo	40 weeks	TZP 5 mg QW 121 TZP 10 mg QW 121 TZP 15 mg QW 121 Placebo QW 115 Total 478 (1:1:1:1)
I8F-MC-GPGL (Add-on to metformin)	M or F, ≥18 years of age with T2DM, on stable metformin (≥1500 mg/day) for ≥3 months prior to study entry; HbA1c of 7.0%-10.5%, inclusive; stable weight ≥3 months prior to study entry; eGFR ≥45 mL/min/1.73 m^2; BMI ≥25 kg/m^2	On stable metformin (≥1500 mg/day) for ≥3 months prior to study entry	Randomized, open-label[b], active controlled	Semaglutide 1.0 mg	40 weeks	TZP 5 mg QW 470 TZP 10 mg QW 469 TZP 15 mg QW 470 Sema 1.0 mg QW 469 Total 1878 (1:1:1:1)

Study	Patient Population (Key Inclusion Criteria)	Background Therapy	Study Design	Control/ Comparator	Treatment Duration[a]	No. of Patients Treated with Study Drug by Treatment (Randomization Ratio)
I8F-MC-GPGH (Add-on to 1 or 2 OAMs)	M or F, ≥18 years of age with T2DM, insulin naive and on stable metformin (≥1500 mg/day) ± SGLT-2i for ≥3 months prior to study entry; HbA1c of 7.0%-10.5%, inclusive; stable weight ≥3 months prior to study entry; eGFR ≥45 mL/min/1.73 m^2; BMI ≥25 kg/m^2	Insulin naive and on stable metformin (≥1500 mg/day) ± SGLT-2i for ≥3 months prior to study entry	Randomized, open-label, active-controlled	Insulin degludec[c]	52 weeks	TZP 5 mg QW 358 TZP 10 mg QW 360 TZP 15 mg QW 359 Insulin degludec QD 360 Total 1437 (1:1:1:1)
I8F-MC-GPGM (Add-on to 1 to 3 OAMs)	M or F, ≥18 years of age with T2DM, on stable treatment of 1-3 OAMs (metformin, SGLT-2i, SU) ≥3 months prior to study entry with increased risk of CV events; HbA1c of 7.5%-10.5%, inclusive; eGFR ≥45 mL/min/1.73 m^2; stable weight ≥3 months prior to study entry; BMI ≥25 kg/m^2	On stable treatment with 1-3 OAMs (metformin, SU, SGLT2i) for ≥3 months prior to study entry	Randomized, open-label, active-controlled	Insulin glargin[d]	52-104 weeks	TZP 5 mg QW 329 TZP 10 mg QW 328 TZP 15 mg QW 338 Insulin glargine QD 1000 Total 1995 (1:1:1:3)
I8F-MC-GPGI (Add-on to basal insulin)	M or F, ≥18 years of age with T2DM, on stable doses of once-daily insulin glargine (>0.25 U/kg/day or >20 U/day) ± metformin (≥1500 mg/day) for ≥3 months prior to study entry; HbA1c of 7.0%-10.5%, inclusive; stable weight ≥3 months prior to study entry; eGFR ≥30 mL/min/1.73 m^2 or ≥45 mL/min/1.73 m^2 (for patients on metformin); BMI ≥23 kg/m^2	Insulin glargine ± metformin for ≥3 months prior to study entry'	Randomized, double-blind, placebo-controlled	Placebo	40 weeks	TZP 5 mg QW 116 TZP 10 mg QW 119 TZP 15 mg QW 120 Placebo 120 Total 475 (1:1:1:1)

Reference ID: 4972437

Study	Patient Population (Key Inclusion Criteria)	Background Therapy	Study Design	Control/ Comparator	Treatment Duration[a]	No. of Patients Treated with Study Drug by Treatment (Randomization Ratio)
Regional (Japan) Phase 3 Study						
I8F-JE-GPGO	M or F, ≥20 years of age with T2DM, OAM naive or have discontinued OAM monotherapy; HbA1c of ≥7.0% to ≤ 10.0%, inclusive (OAM naive), 6.5%-9.0% at Visit 1 and 7.0%-10.0% at Visit 2 for those discontinuing OAM monotherapy; stable weight ≥3 months prior to study entry; eGFR ≥30 mL/min/1.73 m^2 (or lower than the country-specific threshold for	No OAM ≥2 months prior to study entry	Randomized, double- blind, active-controlled	Dulaglutide 0.75 mg	52 weeks	TZP 5 mg QW 159 TZP 10 mg QW 158 TZP 15 mg QW 160 Dula 0.75 mg QW 159 Total 636 (1:1:1:1)

Abbreviations: BMI = body mass index; CV = cardiovascular; Dula = dulaglutide; eGFR = estimated glomerular filtration rate; F = female; HbA1c = glycosylated hemoglobin A1c; M = male; OAM = oral antihyperglycemic medication; QD = once-daily; QW = once-weekly; SU = sulfonylurea; SGLT2i = sodium/glucose co-transporter 2 inhibitor; T2DM = type 2 diabetes mellitus; TZP = tirzepatide.
a All studies included a 4-week follow-up period after the treatment period.
b Within the tirzepatide arms, the dose of tirzepatide was double-blinded.
c Starting dose of 10 IU/day, titrated to FBG <90 mg/dL, following a TTT algorithm.
d Starting dose of 10 IU/day, titrated to FBG <100 mg/dL, following a TTT algorithm.
Study GPGO was not included in the interim CV meta-analysis because the database lock for this study did not occur prior to the interim analysis.

Source: From applicant's multistudy-cv-meta-analysis-t2dm.pdf document (Table 4.1, page 15/394).

Reference ID: 4972437

Of the 7 trials in Table 4, the Phase 2 trial was the shortest with a 26-week duration. This trial was also the only trial that included both placebo and an active-control. Trials GPGI, GPGK and GPGL were 40 weeks in duration, and Trials GPGH, GPGM and GPGO were of at least 52 weeks in duration. Active-controlled trials GPGH, GPGL and GPGM were open-label; Trial GPGO, conducted entirely in Japan was the only active-controlled trial that was double-blind. Trial GPGM included patients with increased CV risk.

3.3.1.2 Treatment Regimens in the Trials

Test drug

Tirzepatide 1mg, 5 mg, 10 mg and 15 mg, all QW

Comparator Drug

Placebo, dulaglutide 0.75 mg QW, dulaglutide 1.5 mg QW, semaglutide 1.0 QW, insulin degludec QD and insulin glargine QD.

3.3.1.3 Endpoints

The sponsor state that unless otherwise specified only CV events that occurred during the treatment period and the 30-day safety follow-up period that were positively adjudicated by the CEC were included in the CV meta-analysis.

The **primary endpoint** is the time-to-event based on CEC-confirmed MACE-4.
Time-to-event is defined as number of days between the date of first dose and

- the onset date of the event plus 1 day if the patient experienced any of the component events in MACE-4 on or before the 30-day safety follow-up visit or
- =the censoring date plus 1 day if the patient does not experience a component event of MACE-4 on or before 30-day safety follow-up visit or early study termination, if applicable.

MACE-4 occurring after the 30-day safety follow-up visit or after withdrawal of consent would not be included in the primary analysis.
Time-to-event analyses were to be performed for all-cause mortality and other additional endpoints (MACE-3, CV death, composite of MACE-3 or hospitalization for heart failure, all myocardial infarction, all stroke, hospitalization for unstable angina, and hospitalization for heart failure) and the hazard ratio and confidence interval from a Cox proportional hazards regression model, and Kaplan-Meier plots were to be generated if the number of events ≥ 10.

3.3.1.4 Sample Size Calculation

Trial GPGM was to contribute the vast majority of events for the meta-analysis, approximately 110 of the 133 MACE-4 events anticipated at the time of the final meta-analysis were expected to be from Trial GPGM.

The applicant assessed the expected number of MACE-4 events in the pooled tirzepatide arms versus the pooled comparator arms across all Phase 3 trials in order to show that if there were no excess risk with tirzepatide then the meta-analysis would have at least 90% power to show that upper bound of the CI for the hazard ratio would be less than 1.8.

3.3.1.5 Clinical Events Adjudication

The Cleveland Clinic Coordinating Center for Clinical Research (C5R) Clinical Events Adjudication (CEC) group was responsible for adjudicating clinical events in Phase 3 trials in a blinded, consistent, and unbiased manner across the studies included in the meta-analysis. The C5R CEC created and maintained the CEC Charter and collaborated with the applicant to adjudicate and classify:

- All deaths – including CV and non-CV deaths
- Acute coronary syndromes: MI, HUA
- Cerebrovascular events: stroke, transient ischemic attack (TIA)
- Hospitalization for heart failure
- Coronary interventions (such as coronary artery bypass graft or percutaneous coronary intervention)

3.3.2 Statistical Methodology

3.3.2.1 Analysis Populations

All analyses were to be conducted on the modified-intent-to-treat (mITT) population, which consists of all randomized patients receiving at least 1 treatment dose according to the treatment to which they were assigned. Only CV events that occurred during the treatment period and the 30-day safety follow-up period that were positively adjudicated by the CEC were included in the CV meta-analysis.

3.3.2.2 Analysis Methods for Safety Endpoints

Primary MACE-4 endpoint:

The primary objective was to demonstrate that tirzepatide was not associated with an unacceptably high risk for MACE-4 in patients with T2DM. This objective was to be evaluated by comparing the distribution of time from first dose to the first occurrence of MACE-4, for patients receiving any dose of QW tirzepatide (pooled tirzepatide group) to that in patients administered comparators, placebo, or active control (pooled control group). The primary objective was to be considered met if the upper bound of the 2-sided 95% confidence interval (CI) for the hazard ratio (HR) (pooled tirzepatide versus pooled control) from the meta-analysis was <1.8.

The **primary analysis** for the primary endpoint estimates a hazard ratio for treatment (pooled tirzepatide arms) versus control (pooled control arms) using a Cox proportional hazards model with treatment (pooled tirzepatide arms, pooled control arms) as a fixed effect and stratified by study-level CV risk:

- Stratum 1: Study GPGM (high CV risk patient population with longer follow-up) and
- Stratum 2: Studies GPGB, GPGH, GPGI, GPGK, GPGL, and 18F-JE-GPGO (lower CV risk population),

was to be used to estimate the hazard ratio and associated confidence interval using the mITT population.

Reference ID: 4972437

3.3.2.3 Interim Analyses and Multiple Testing Plan

A single interim analysis of the CV meta-analysis was to be conducted when all of the following conditions were met:

1. At least 100 patients had reached primary MACE-4 endpoint confirmed by the CEC across all the trials in Table 4.1 above, including Study GPGM.
2. All global trials (Studies GPGK, GPGI, GPGH, and GPGL) described in Table 4.1 had achieved their database lock.
3. All patients in Study GPGM who had not discontinued the study before 12 months had completed the 12-month primary endpoint assessment.
4. At least 300 patients in Study GPGM had reached 18 months or longer of exposure to tirzepatide.

Note: If 133 patients had reached primary MACE-4 endpoint confirmed by the CEC across the trials at the time conditions 2, 3, and 4 were fulfilled, no interim analysis was to be conducted. In this setting, the final analysis for all endpoints would be conducted utilizing alpha=0.05.

The applicant proposed to use a 2-sided alpha=0.01 at the interim analysis with CEC confirmed MACE-4 in 100 patients. That is, if the 99% CI for HR is (combined tirzepatide versus combined control) <1.8, then it would be concluded that tirzepatide would not result in an unacceptable increase in CV risk. If the upper bound of 99% CI ≥1.8, Study GPGM was to be continued until the accrual of approximately 133 patients that experience MACE-4 endpoints across the tirzepatide development program, and the final analysis would be conducted comparing the upper bound of 95.2% CI of HR (combined tirzepatide versus combined control) to 1.8. Alpha levels for the interim and final CV meta-analysis were to be calculated using EAST software and corresponded to the Hwang, Shih, DeCani spending function with Gamma=-6.6.

If the number of patients with CEC confirmed MACE-4 at the interim analysis exceeded 100, alpha level for the interim analysis was to be determined based on Hwang, Shih, DeCani with Gamma=-6.6 and the information fraction based on the total of 133 endpoints.

If the number of patients with MACE-4 endpoints at the end of the study differed from 133, the alpha level available for the final analysis would be recalculated based on the alpha spent at the interim, the number of patients with MACE-4 at the interim analysis, and the number of patients with MACE-4 at the final analysis.

The applicant did not specify a multiple testing plan that included endpoints other than the MACE-4 endpoint. The all-cause mortality endpoint was evaluated at the 5% level of significance using the final data.

3.3.3 Patient Disposition, Demographics and Baseline Characteristics

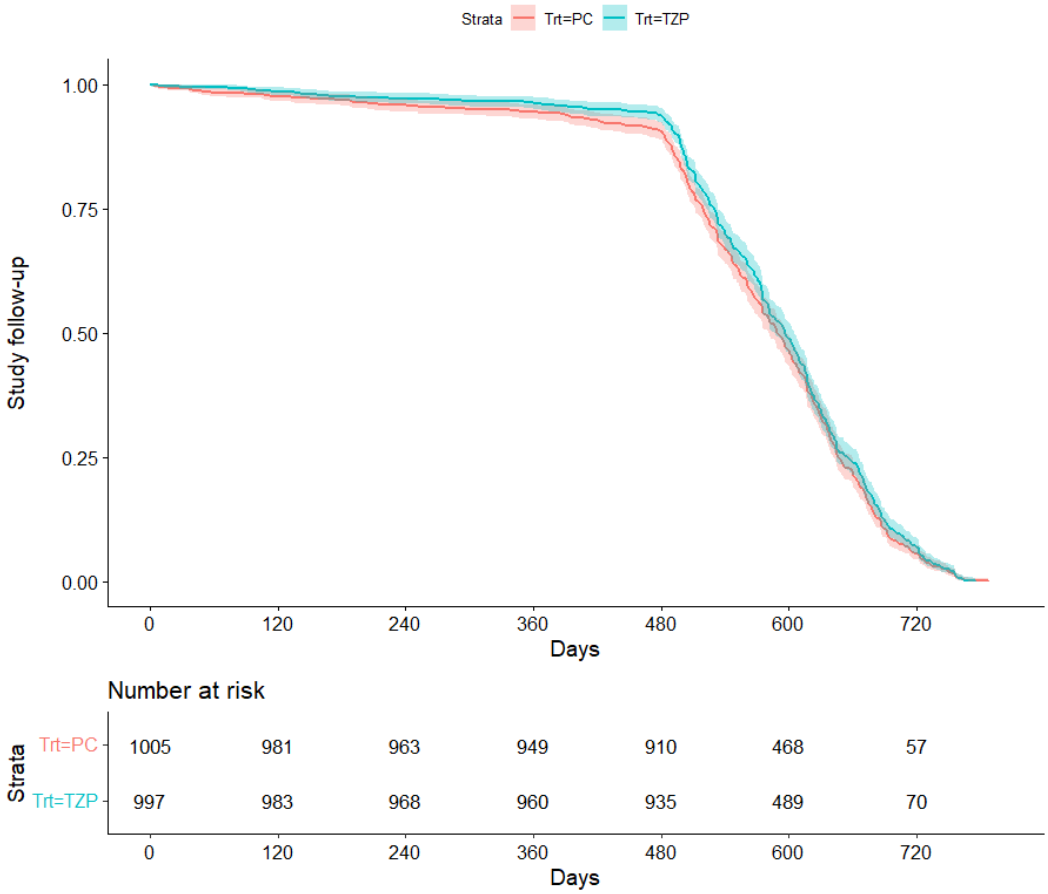
The studies included in the CV meta-analysis collectively screened 10185 patients and randomized 7232 patients – 4894 in the tirzepatide arms and 2338 in the pooled comparator arm. As indicated in Table 4 above, randomization ratios varied for tirzepatide arms versus control arms in these trials (1:1, 4:2 and 3:1) – all except Trial GPGB were Phase 3 trials.

Reference ID: 4972437

3.3.3.1 Patient Follow-up and Disposition

Figure 1 shows study follow-up by pooled tizepatide (green) and pooled comparator (orange) arms for Trial GPGM and Figure 2 shows follow-up for all other trials combined. While follow-up appears similar across treatment arms, the tirzepatide arm appear to have slightly longer follow-up than the pooled comparator arm in Trial GPGM.

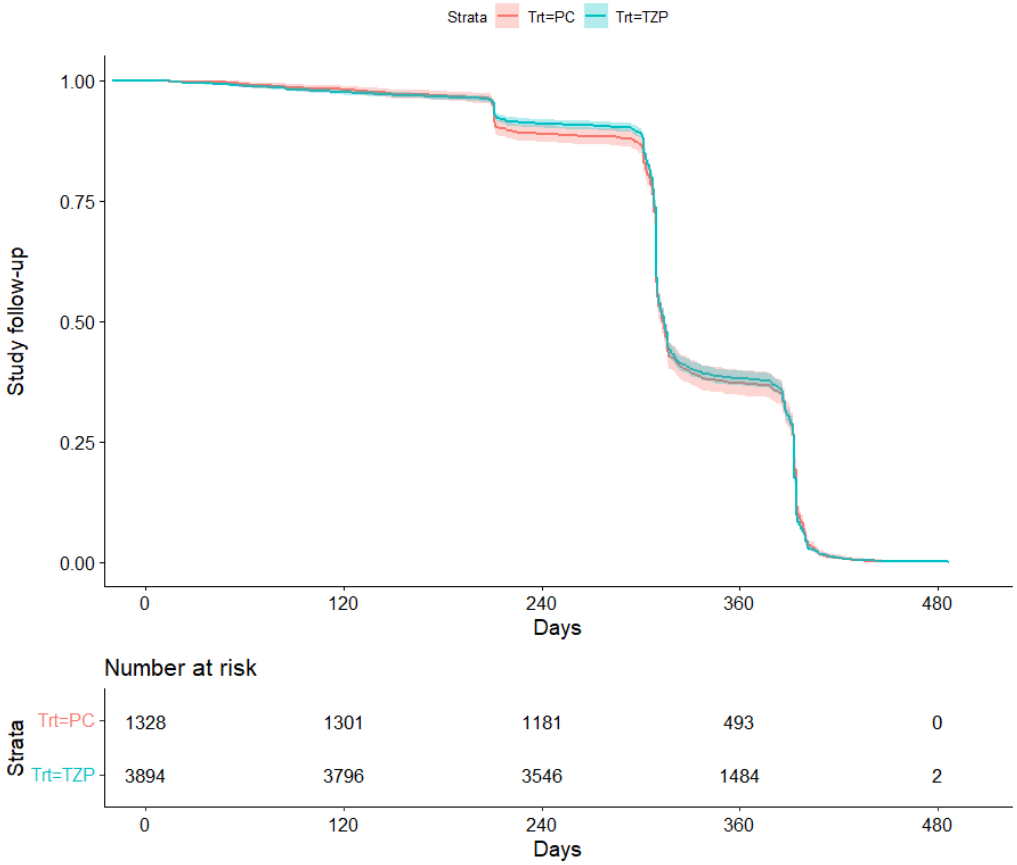
Figure 1: Follow-up by treatment arm for Trial GPGM (ITT)



Source: Created by FDA statistical reviewer

Reference ID: 4972437

Figure 2: Follow-up by treatment arm for trials other than GPGM (ITT)



Source: Created by FDA statistical reviewer

As indicated in Table 5, the mean follow-up time for the pooled comparator arm in Trial GPGM was 1.60 years with a standard deviation of 0.31 years. The mean follow-up time for the pooled comparator arms was smaller at 1.56 years with a standard deviation of 0.36. The mean and standard deviation over all other trials combined was the same for both arm with a mean of 0.9 years and a standard deviation of 0.19 years.

Table 5. Subject follow-up in years by stratum and treatment arm (ITT)

	Pooled tirzepatide	**Pooled Comparator**
Trial GPGM	1.60 (0.31)	1.56 (0.36)
All other trials combined	0.90 (0.19)	0.90 (0.19)

Source: Created by the FDA statistical reviewer.

9.32% (218/2338) of randomized patients on the pooled comparator arm did not complete the study, and 6.70% (328/4894) on the pooled tirzepatide arms did not, i.e., collectively a slightly larger fraction of subjects on the control arms discontinued the study. Table 6 gives study discontinuation reasons by treatment arm by trial for the randomized patients. Although there are numerical differences in the fractions of discontinuations across arms, in most cases the counts are small, and differences are likely by chance. The largest counts are due to "withdrawal by

Reference ID: 4972437

subject" – there are some differences in discontinuations across arms for both double-blind and open-label studies (highlighted in gray in the table) but the direction of these differences is not consistent across studies. In addition, since "withdrawal by subject" can include a variety of different reasons, no clear interpretation of potential differences is possible.

Table 6: Discontinuation reasons by Study (ITT)

Study ID	Treatment arm (N)	Discontinuation Reason N (%)							
		Adverse Event	Death	Lost to follow-up	Physician Decision	Protocol Deviation	Screen failure	Withdrawal by subject	Misc*
GPGB	PC (105)	2 (1.90)	1 (0.95)	3 (2.86)	0 (0.00)	0 (0.00)	0 (0.00)	4 (3.81)	1 (0.95)
	TZP (213)	4 (1.88)	0 (0.00)	9 (4.23)	0 (0.00)	0 (0.00)	1 (0.47)	13 (6.10)	1 (0.47)
GPGH	PC (365)	1 (0.27)	1 (0.27)	5 (1.37)	2 (0.55)	0 (0.00)	0 (0.00)	22 (6.03)	3 (0.82)
	TZP (1079)	15 (1.39)	4 (0.37)	22 (2.04)	4 (0.37)	2 (0.19)	0 (0.00)	32 (2.97)	6 (0.56)
GPGI	PC (120)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.83)	0 (0.00)	2 (1.67)	0 (0.00)
	TZP (355)	0 (0.00)	0 (0.00)	1 (0.28)	0 (0.00)	3 (0.85)	0 (0.00)	12 (3.38)	0 (0.00)
GPGK	PC (115)	1 (0.87)	1 (0.87)	5 (4.35)	2 (1.74)	1 (0.87)	0 (0.00)	6 (5.22)	0 (0.00)
	TZP (363)	5 (1.38)	0 (0.00)	8 (2.20)	1 (0.28)	1 (0.28)	0 (0.00)	18 (4.96)	1 (0.28)
GPGL	PC (469)	3 (0.64)	1 (0.21)	12 (2.56)	4 (0.85)	0 (0.00)	0 (0.00)	4 (0.85)	2 (0.43)
	TZP (1410)	6 (0.43)	12 (0.85)	19 (1.35)	2 (0.14)	1 (0.07)	0 (0.00)	22 (1.56)	8 (0.57)
GPGM	PC (1009)	10 (0.99)	35 (3.47)	22 (2.18)	6 (0.59)	2 (0.20)	0 (0.00)	40 (3.96)	8 (0.79)
	TZP (997)	8 (0.80)	25 (2.51)	15 (1.50)	2 (0.20)	3 (0.3)	1 (0.10)	17 (1.71)	7 (0.70)
GPGO	PC (159)	4 (2.52)	0 (0.00)	1 (0.63)	1 (0.63)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	TZP (477)	4 (0.84)	0 (0.00)	0 (0.00)	1 (0.21)	0 (0.00)	0 (0.00)	10 (2.10)	1 (0.21)

*Misc: Includes Pregnancy, Study terminated by sponsor and Other
Rows highlighted in gray are open label studies.
Source: Created by FDA statistical reviewer.

3.3.3.2 Demographics and Baseline Characteristics

Baseline demographic characteristics for all randomized subjects by treatment arm are presented in Table 8 and baseline clinical characteristics are presented in Table 9. These tables indicate that the treatment arms are reasonably balanced in terms of both demographic and clinical characteristics at baseline.

Reference ID: 4972437

Approximately 43% of subjects were female, the mean age of subjects was approximately 59 years, 73% of subjects were of White race, 7% were American Indian or Alaskan Native, 42% identified as Hispanic or Latino, 22% were US subjects, and Central/South America and Mexico constituted 33% of patients randomized. Mean BMI at baseline was 33 kg/m2 in both arms, mean duration of T2DM was a little over 9 years, 82-83% reported no tobacco use at baseline, eGFR at baseline was 89 ml/min/1.73m^2, mean UACR was around 85-86 g/kg and mean HBA1c was around 8%in both arms.

Table 7: Baseline Demographic Characteristics (ITT)

N (%)	**Pooled tirzepatide (N=4894)**	**Pooled Comparator (N=2338)**
Sex		
Female	2166 (44.26)	963 (41.19)
Male	2728 (55.74)	1375 (58.81)
Age in years		
Mean ± SD	58.25 (10.35)	60.01 (10.23)
Median (IQ range)	59.00 (15.00)	61.00 (13.00)
Race		
White	3513 (71.78)	1788 (76.48)
Black or African-American	186 (3.80)	71 (3.04)
Asian	792 (16.18)	275 (11.76)
Native Hawaiian or other Pacific Islander	8 (0.16)	4 (0.17)
American Indian or Alaskan Native	356 (7.27)	165 (7.06)
Multiracial	36 (0.74)	32 (1.37)
Ethnic Group* **(**Hisp/Latino**)**	2041 (41.70)	1021 (43.67)
U.S.A.	1088 (22.23)	506 (21.64)
Region		
Asia (excluding Japan)	205 (4.19)	86 (3.68)
Central/South America and Mexico	1592 (32.53)	830 (35.50)
EU/United Kingdom/Ukraine	1248 (25.50)	607 (25.96)
Japan	605 (12.36)	202 (8.64)
North America	1156 (23.62)	548 (23.44)
Rest of the world	88 (1.80)	65 (2.78)

*: Ethnicity was either not reported or unknown for 636 and 218 subjects respectively in the pooled tirzepatide and comparator arms.

Source: Created by FDA statistical reviewer.

Reference ID: 4972437

Table 8: Baseline Clinical Characteristics (ITT)

	Pooled tirzepatide N=4894	**Pooled Comparator N=2338**
BMI[a] (kg/m^2)		
Mean ± SD	32.87 ± 6.35	32.63 ± 6.13
Median (IQ range)	31.79 (4.38)	31.53 (7.78)
Duration of Diabetes[b] (years)		
Mean ± SD	9.00 ± 6.89	9.81 ±7.24
Median (IQ range)	7.55 (8.46)	8.42 (9.43)
Tobacco Use[c] N(%)		
No	4070 (83.16)	1916 (81.95)
Yes	824 (16.84)	421 (18.01)
eGFR[d] (ml/min/1.73m^2)		
Mean ± SD	89.82 ± 19.19	87.4 ± 19.79
Median (IQ range)	93.00 (26)	91.00 (25)
UACR[e] (g/kg)		
Mean ± SD	85.23 ± 354.32	86.00 ± 363.95
Median (IQ range)	11.00 (32)	12.00 (36.59)
Baseline A1C[f] (%)		
Mean ± SD	8.28 ± 0.96	8.32 ± 0.91
Median (IQ range)	8.10 (1.3)	8.20 (1.3)

a: These BMI calculations excluded 16 missing values
b: These duration of diabetes calculations excluded 1 missing value.
c: These tobacco-use tabulations excluded 1 missing value.
d: These eGFR calculations excluded 19 missing values.
e: These UACR calculations excluded 68 missing values.
f: These A1C calculations excluded 18 missing values.

Source: Created by FDA statistical reviewer.

3.3.4 Results and Conclusions

3.3.4.1 Primary Safety Endpoint Analyses

The pre-specified primary analysis for the MACE-4 endpoint was a Cox proportional hazards model that included a fixed treatment effect and was stratified by study-level CV risk. This was analyzed on the mITT population and only adjudicated MACE-4 events are considered in the analyses for this endpoint presented in this review.

A pre-specified interim analysis was conducted for the CV meta-analysis. At the interim analysis 116 events were observed for the MACE-4 endpoint. The applicant stated in its response to an FDA IR that the database lock date for the interim analysis was February 24, 2021 and that the database lock for Trial GPGO had not occurred at this time and hence it was not included in this interim analysis.

Reference ID: 4972437

Table 10 shows the number of MACE-4 events observed by trial and treatment arm. The majority of events (86/116) observed were from Trial GPGM that enrolled high-risk patients; this trial had an overall MACE-4 incidence rate of 3.22 events per 100 patient-years. GPGM was an open-label trial with active comparator insulin glargine. All other trials had incidence rates of less than or equal to 1.13 events per 100 patient-years. The next largest incidence rate was contributed by Trial GPGB with 1.13 MACE-4 events per 100 patient-years. Among Phase 3 trials, the next highest incidence rate (0.89 events per 100 patient-years) was observed in Trial GPGL. This trial was also an open-label trial with semaglutide 1 mg as the active comparator. For all studies except GPGL the percentage of subjects with MACE-4 events was higher in the comparator arm than in the tirzepatide arm. Trial GPGL, which had semaglutide 1 mg as the active comparator, had an incidence rate of 1.10 per 100 patient-years in the pooled tirzepatide arms and 0.25 per 100 patient-years in the semaglutide 1 mg arm.

Table 9: Patients with MACE-4 events by Trial (mITT, IOMA)*

Trial	Tirzepatide	Pooled Comparator	Total
	Num. Events/Num. Pts. IR/100 PY)		
GPGM [a]	37/995 (2.74)	49/1000 (3.70)	86/1995 (3.22)
GPGB [b]	1/211 (0.85)	1/105 (1.71)	2/316 (1.13)
GPGH [c]	7/1077 (0.62)	3/360 (0.80)	10/1437 (0.67)
GPGI	2/355 (0.69)	1/120 (1.00)	3/475 (0.77)
GPGK	0/363 (0.00)	1/115 (1.08)	1/478 (0.26)
GPGL [d]	13/1409 (1.10)	1/469 (0.25)	14/1878 (0.89)
Total	**60/4410 (1.38)**	**56/2169 (2.39)**	**116/6579 (1.73)**

*: mITT: All randomized subjects who took at least one dose of study drug; IOMA: Interim look of the meta-analysis
a: Insulin glargine was the active comparator for this trial.
b: Dulaglutide 1.5 mg was the active comparator for this trial.
c: Insulin degludec was the active comparator for this trial.
d: Semaglutide 1 mg was the active comparator for this trial.

Source: Created by FDA statistical reviewer.

The pre-specified primary MACE-4 analysis, that compares pooled tirzepatide doses with pooled comparator is presented in Table 11. The estimated hazard ratio (HR) for pooled tirzepatide doses to pooled comparator was **0.81** and the corresponding 97.85% confidence interval (CI) **(0.52, 1.26)**[4] has an upper bound less than 1.8, thus meeting the risk margin specified in the meta-analysis statistical analysis plan. The Kaplan-Meier cumulative incidence plot[5] in Figure 4 shows that the estimated cumulative incidence for MACE-4 for the combined tirzepatide arm is below that for the pooled comparator arm although, as noted above, the confidence interval for the hazard ratio does not indicate a significant difference between hazards for pooled tirzepatide and pooled comparator.

[4] A Cox PH analysis that stratified by individual studies resulted in a hazard ratio of 0.82 with an associated 97.85% CI of (0.53, 1.26).

[5] The applicant presents what it refers to as "adjusted Kaplan-Meier plot" estimated by weighting with inverse probability of randomization for treatment within strata. These differ slightly from the Kaplan-Meier plots in this review.

Reference ID: 4972437

Also presented in Table 11 are hazard ratios by tirzepatide dose. We see that there is a numerical decrease in incidence rates with increasing tirzepatide dose, i.e., the 1 mg dose has the largest observed incidence rate and the 15 mg dose has the smallest. A similar pattern is observed in the HR estimates; all 97.85% confidence intervals for tirzepatide doses versus pooled comparator include 1, i.e., no dose indicates a nominally significantly different hazard from that of the pooled comparator. Figure 5 shows that tirzepatide 5, 10 and 15 mg doses and pooled comparator show similar incidence in earlier weeks with visible separation occurring in later weeks – pooled comparator curve is higher, indicating higher incidence. Only Trial GPGB included the 1 mg dose; with a single event in this dose arm, there was insufficient information to obtain reliable HR estimates for this dose.

Table 10: Analysis of primary MACE-4 endpoint in CV meta-analysis by tirzepatide dose (mITT; IOMA)*

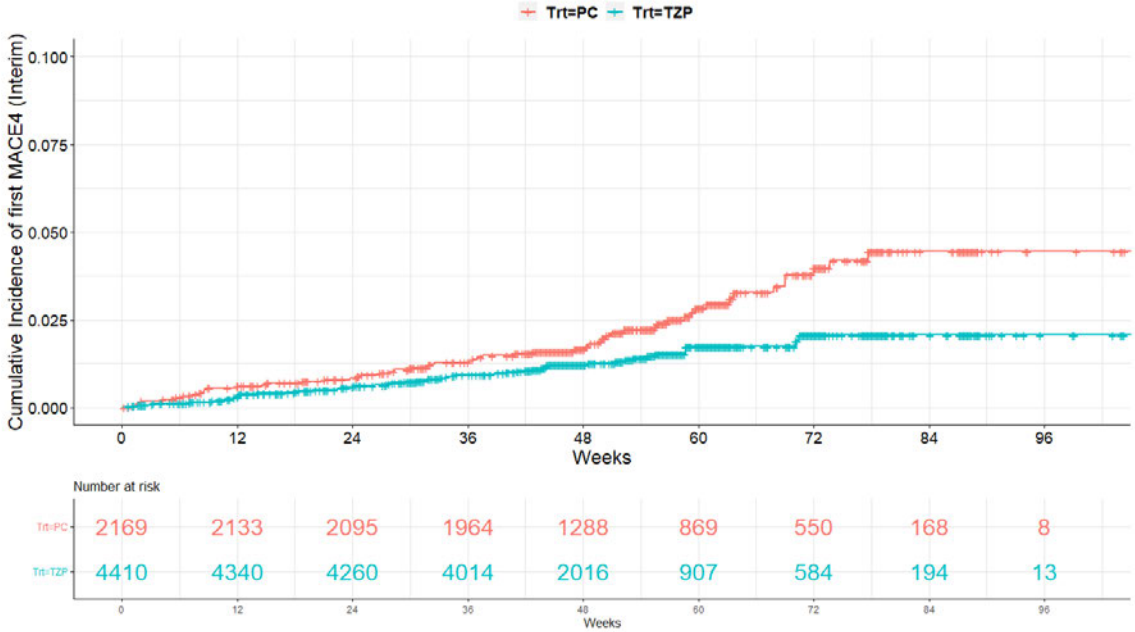
	Tirzepatide					**Pooled Comparator**
	15 mg N= 1461	**10 mg N=1448**	**5 mg N= 1449**	**1 mg N=52**	**All N= 4410**	**N=2169**
MACE-4 patient years of follow-up	1457.19	1435.64	1428.04	28.59	4349.45	2344.46
MACE-4 events (rate per 100 PY)	16 (1.10)	18 (1.25)	25 (1.75)	1 (3.50)	60 (1.38)	56 (2.39)
HR (97.85% CI)§	0.64 (0.33, 1.24)	0.74 (0.40, 1.39)	1.04 (0.59, 1.82)	5.01 (0.46, 54.75)	0.81 (0.52, 1.26)	-

* **mITT**: All randomized subjects who took at least one dose of study drug; **IOMA**: Interim look of the meta-analysis
‡ Hazard ratios for pooled tirzepatide doses computed using a Cox proportional hazards model with fixed effects for treatment and stratified by study-level CV Risk.
§ The confidence level of 97.85 was based on the 116 events observed at interim and the pre-specified Hwang, Shih and DeCani spending function with Gamma=-6.6.

Source: Created by FDA statistical reviewer using adtte provided by applicant in February 2022.

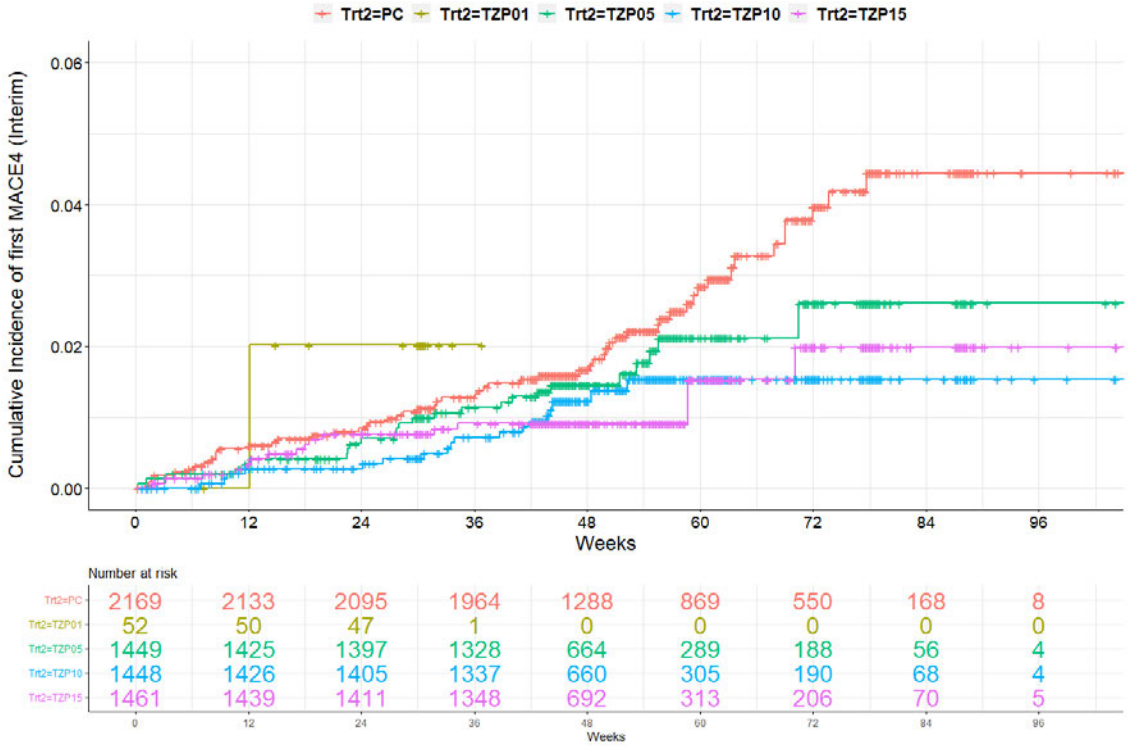
Reference ID: 4972437

Figure 3: Kaplan-Meier-based cumulative incidence plots for MACE-4 by treatment arm (mITT, IOMA)*



* **mITT**: All randomized subjects who took at least one dose of study drug; **IOMA**: Interim look of the meta-analysis
Source: Created by FDA statistical reviewer.

Figure 4: Kaplan-Meier-based cumulative incidence plots for MACE-4 by treatment arm and dose (mITT, IOMA)*



* **mITT**: All randomized subjects who took at least one dose of study drug; **IOMA**: Interim look of the meta-analysis
Source: Created by FDA statistical reviewer.

Reference ID: 4972437

Reviewer Comments:

- *Of the 6 studies included in the interim look of the meta-analysis for the primary MACE-4 endpoint, 3 studies – GPGH, GPGL and GPGM – were open label studies. These studies resulted in most of the MACE-4 events (86+14+10=110 of 116 total MACE-4 events observed at the interim look). The lack of blinding in these open-label studies has the potential to introduce bias in study conduct and results and should be considered when interpreting study results.*
- *Different active comparators were used in the collection of studies in the interim meta-analysis. The open-label study GPGM, that contributed the most events (86/116), compared tirzepatide to insulin glargine and resulted in a lower incidence rate per 100 patient-years in the tirzepatide arm than in the pooled comparator arm (2.74 compared to 3.70). Study GPGL – an open-label study that contributed the next highest number of events to the interim analysis (14/116) – had a higher incidence rate in the tirzepatide arm compared to the semaglutide 1 mg arm (1.10 versus 0.25). There were insufficient data in the 3 randomized, double-blind, placebo-controlled trials – GPGB, GPGI and GPGK— which contributed a total of 6 MACE-4 events to this interim analysis. The comparators used and number of events observed in the different studies in the meta-analysis should be considered while clinically interpreting the hazard ratio estimates.*

3.3.4.2 Supporting analyses for Primary MACE endpoint

We present in this sub-section additional analyses that explore the components of the MACE-4 endpoint at the interim analysis, and the analyses of the MACE-4 endpoint using end-of-meta-analysis (EOMA)/complete data. The analyses of some components of the MACE-4 endpoint were pre-specified by the sponsor under "Other secondary endpoints", and some on-treatment analyses were proposed as additional analyses. None of these analyses were covered under a multiple testing plan and are hence considered exploratory.

3.3.4.2.1 MACE-4 components at interim analysis

Analyses of the components of the primary MACE-4 endpoint (CV death, MI, stroke and hospitalization for unstable angina) for the interim analysis are presented in Table 12. Only CV events that occurred during the treatment period and the 30-day safety follow-up period that were positively adjudicated by the CEC were included in the CV meta-analysis.
The incidence rate per 100 patient years for the pooled comparator arm was numerically larger for all of the MACE-4 components and approximately double that in the pooled tirzepatide arm for all except the CV death component. Although hazard ratios estimated were numerically less than 1 for all of the MACE-4 components, the associated 97.85% confidence intervals include the null value of 1.

Table 11: Description of components of primary MACE-4 endpoint in CV meta-analysis (mITT; IOMA)*

	All Tirzepatide N=4410 PY=4349.45§	Pooled Comparator N=2169 PY=2344.46§	Hazard Ratio‡ (97.85% CI)
	Num. Events (IR per 100 PYs)		

Reference ID: 4972437

Primary MACE-4 events	60 (1.38)	56 (2.39)	0.81 (0.52, 1.26)
Components†:			
CV death	20 (0.46)	15 (0.63)	0.99 (0.44, 2.22)
MI	23 (0.53)	26 (1.10)	0.65 (0.33, 1.27)
Stroke	13 (0.30)	14 (0.59)	0.76 (0.31, 1.88)
Hospitalization for Unstable Angina	5 (0.11)	6 (0.25)	0.75 (0.18, 3.08)

* **mITT**: All randomized subjects who took at least one dose of study drug; **IOMA**: Interim look of the meta-analysis
§: These patient years are for the primary MACE-4 events; patient years for components differed slightly from these.
†: Analyses for components capture all CV deaths, all MIs (fatal and non-fatal) and all strokes (fatal and non-fatal). Some subjects experienced multiple events, hence totals of component events exceed number of primary MACE-4 events, which only considers first event for each patient.
‡ Hazard ratios for pooled tirzepatide doses computed using a Cox proportional hazards model with fixed effect for treatment and stratified by study-level CV risk (GPGM versus all others). The confidence level of 97.85 was based on the 116 events observed at interim and the pre-specified Hwang, Shih and DeCani spending function with Gamma=-6.6.

Source: Created by FDA statistical reviewer.

3.3.4.2.2 Analyses of MACE-4 at End-of-study

Table 12 descriptively presents MACE-4 event counts, patient numbers and incidence rates per 100 patient-years by study and treatment arm. Trial GPGO was not included in the interim analysis for the MACE-4 endpoint, this study is included in the end-of-meta-analysis (EOMA) analyses or complete data analyses. Since studies other than GPGO and GPGM had already been completed at the time of the interim analysis, we see no changes to counts and incidence rates for these studies when compared to Table 9. In both Trials GPGM and GPGO the incidence rate of MACE-4 events in the pooled comparator arm was greater than that in the pooled tirzepatide arm. As noted before, this is true in all trials except GPGL.

Table 12: Patients with MACE-4 events by Trial (mITT, EOMA)*

Trial	Tirzepatide	Pooled Comparator	Total
	Num. Events/Num. Pts. (IR per 100 PY)		
GPGM [a]	47/995 (3.01)	62/1000 (4.05)	**109/1995 (3.52)**
GPGO [b]	2/477 (0.40)	1/159 (0.60)	**3/636 (0.45)**
GPGB [c]	1/211 (0.85)	1/105 (1.71)	**2/316 (1.13)**
GPGH [d]	7/1077 (0.62)	3/360 (0.80)	**10/1437 (0.67)**
GPGI	2/355 (0.69)	1/120 (1.00)	**3/475 (0.77)**
GPGK	0/363 (0.00)	1/115 (1.08)	**1/478 (0.26)**
GPGL [e]	13/1409 (1.10)	1/469 (0.25)	**14/1878 (0.89)**
Total	**72/4887 (1.42)**	**70/2328 (2.57)**	**142/7215 (1.82)**

* **mITT**: All randomized subjects who took at least one dose of study drug; **EOMA**: End-of-meta-analysis
a: Insulin glargine was the active comparator for Trial GPGM.
b: Dulaglutide 0.75 mg was the active comparator for Trial GPGO
c: Dulaglutide 1.5 mg was the active comparator for this trial.
d: Insulin degludec was the active comparator for Trial GPGH.
e: Semaglutide 1 mg was the active comparator for this trial.

Source: Created by FDA statistical reviewer.

Reference ID: 4972437

Table 13 contains the analysis of the MACE-4 endpoint by dose using the complete data at the end of the meta-analysis; each tirzepatide dose as well as the pooled tirzepatide arm is compared with the pooled comparator arm. The exposure adjusted incidence rate on the pooled tirzepatide arm (1.42 MACE-4 events per 100 patient years) is a little over half of that in the pooled comparator arm (2.58 MACE-4 events per 100 patient years). A hazard ratio of 0.80 was estimated for the pooled tirzepatide dose versus the pooled comparator with an associated, unadjusted 95% CI of (0.57, 1.11). The upper bound of this interval is below 1.8 and is consistent with the result of the primary MACE-4 analysis at interim. Figure 4 plots the associated cumulative incidence by pooled treatment arms.
We see in Table 13 numerically decreasing hazard ratios for the tirzepatide arms as the dose increases – the 1 mg arm has an observed incidence rate of 3.5 MACE-4 events per 100 patient-years, while the 15 mg arm has an observed incidence rate of 1.06 events per 100 patient years. Whereas the unadjusted 95% CIs for the hazard ratios for other doses include the null value of 1, the CI for the 15 mg dose is entirely below 1. These exploratory analyses seem to indicate the potential for a protective effect for MACE-4 for the 15 mg dose. A visual separation of the 15 mg dose from the pooled comparator is apparent in the cumulative incidence curve for MACE-4 by treatment arm in Figure 5.

Table 13: Analysis of primary MACE-4 endpoint in CV meta-analysis by tirzepatide dose (mITT; EOMA)*

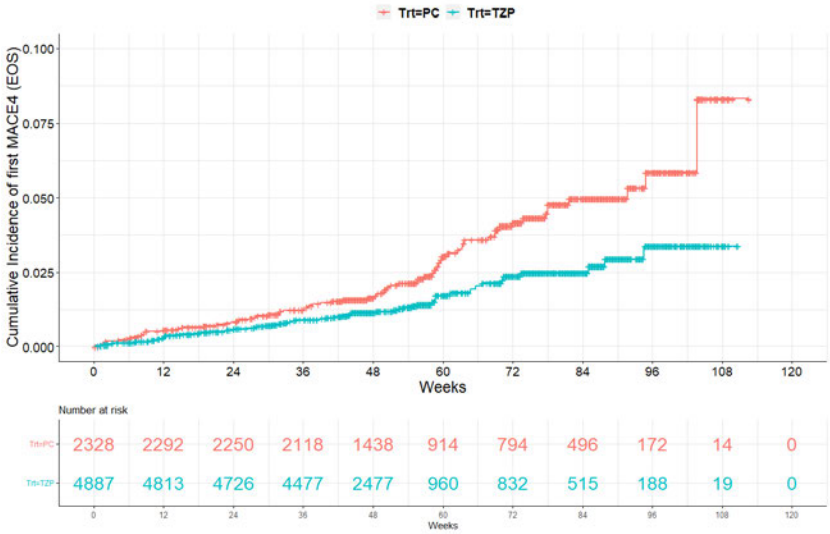
	Tirzepatide					**Pooled Comparator**
	15 mg N=1621	**10 mg N=1606**	**5 mg N= 1608**	**1 mg N=52**	**All N= 4887**	**N=2328**
MACE-4 patient years of follow-up	1700.19	1669.13	1666.53	28.59	5064.45	2717.35
MACE-4 events (IR per 100 PY)	18 (1.06)	25 (1.50)	28 (1.68)	1 (3.50)	72 (1.42)	70 (2.58)
HR‡ (95% CI)	0.59 (0.35, 0.99)	0.84 (0.53, 1.34)	0.95 (0.61, 1.48)	4.73 (0.62, 35.86)	0.80 (0.57, 1.11)	-

* **mITT**: All randomized subjects who took at least one dose of study drug; **EOMA**: End-of-meta-analysis
‡ Hazard ratios for pooled tirzepatide doses computed using a Cox proportional hazards model with fixed effects for treatment and stratified by CV risk (GPGM versus others)

Source: Created by FDA statistical reviewer.

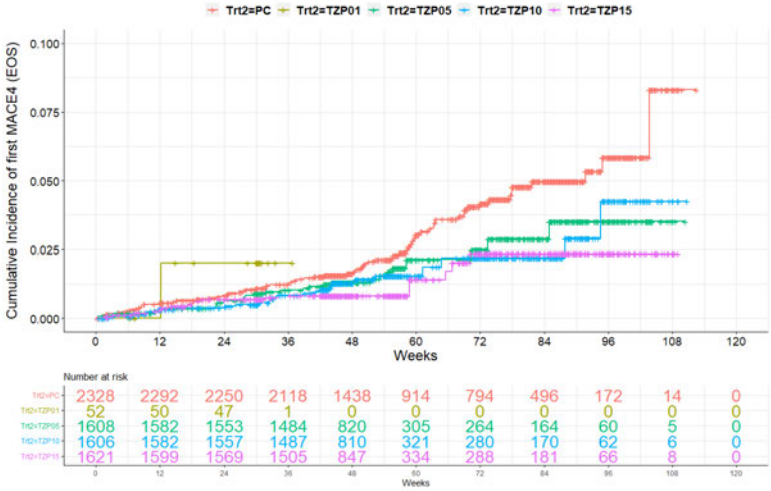
Reference ID: 4972437

Figure 5: Kaplan-Meier-based cumulative incidence plots for MACE-4 by treatment arm (mITT, EOMA)*



* mITT: All randomized subjects who took at least one dose of study drug; EOMA: End-of meta-analysis
Source: Created by FDA statistical reviewer.

Figure 6: Kaplan-Meier-based cumulative incidence plots for MACE-4 by treatment arm and dose (mITT, EOMA)*



* mITT: All randomized subjects who took at least one dose of study drug; EOMA: End-of meta-analysis
Source: Created by FDA statistical reviewer.

3.3.4.2.2.1 Components of MACE-4 at the end of the meta-analysis

Analyses of the components of the primary MACE-4 endpoint (CV death, MI, stroke and hospitalization for unstable angina) at the end of the meta-analysis are presented in Table 15. The incidence rate per 100 patient years for the pooled comparator arm was numerically larger for all of the MACE-4 components. Although hazard ratios estimated were numerically less than 1 for all of the MACE-4 components, the associated 95% confidence intervals include the null value of 1. These results are consistent with the results at the interim analysis.

Reference ID: 4972437

Table 14: Description of MACE-4 endpoint and components (mITT, EOMA)*

	Tirzepatide N=4887 PY=5099.66§ NumEvents (IR/100PY)	Pooled Comparator N=2328 PY=2756.39§ NumEvents (IR/100PY)	Hazard Ratio‡ (95% CI)
MACE-4 (mITT, EOS)	72 (1.42)	70 (2.58)	0.80 (0.57, 1.11)
Components†:			
CV death	25 (0.49)	22 (0.80)	0.90 (0.45, 1.79)
Non-fatal MI	30 (0.59)	30 (1.10)	0.76 (0.41, 1.40)
Non-fatal ischemic stroke	15 (0.29)	15 (0.55)	0.81 (0.34, 1.90)
Hospitalization for Unstable Angina	5 (0.10)	9 (0.33)	0.46 (0.13, 1.71)

* **mITT**: All randomized subjects who took at least one dose of study drug; **IOMA**: Interim look of the meta-analysis
§: These patient years are for the primary MACE-4 events; patient years for components differed from these.
†Analyses capture all components of MACE-4. Some subjects experienced multiple events, hence totals of component events exceeds number of secondary MACE-4 events which only considers first event for each subject.
‡ All Cox proportional hazards analyses used here to compute hazard ratio are stratified by CV risk.

Source: Created by FDA statistical reviewer.

The exploratory analyses presented in this section generally support the conclusion from the interim analysis for the primary MACE-4 endpoint that the pre-specified non-inferiority margin has been met. The exploratory analyses with the complete study data indicate that there is the potential for a dose effect, with the 15 mg dose of tirzepatide potentially being protective for MACE-4. However, the fact that 3 of the 7 studies in this meta-analysis were open-label studies that could be possibly providing biased estimates, and that most of the events were from Trial GPGM in which insulin glargine was the active comparator, need to be factored into the interpretation of these results. These results may not hold for other comparators or in larger randomized studies.

3.3.4.3 All-cause mortality Analyses

The all-cause mortality endpoint was listed as an additional endpoint in the Statistical Analysis Plan (SAP) and the pre-specified primary analysis for the all-cause mortality (ACM) endpoint compared pooled tirzepatide doses with pooled comparator using a Cox proportional hazards model that included a fixed treatment effect and was stratified by study-level CV risk. This analysis was to be carried out on the mITT population based on on-study follow-up. This pre-specified analysis is presented on end-of-meta-analysis data in Table 16 along with comparisons of each dose of tirzepatide with the pooled comparator. A hazard ratio of 0.80 with an associated 95% CI of (0.51, 1.25)[6] is estimated for the comparison of pooled tirzepatide with pooled comparator. This confidence interval covers the null value of 1. Figure 6 shows similar incidence across pooled tirzepatide and comparator arms till about Week 60, after which the cumulative incidence on the pooled comparator arm is numerically higher than that in the pooled tirzepatide

[6] The hazard ratio and 95% CI estimates remained the same when the Cox proportional hazards model was stratified by individual study.

Reference ID: 4972437

arm. No dose trend is evident in either incidence rate estimates, hazard ratio estimates or in Figure 7.

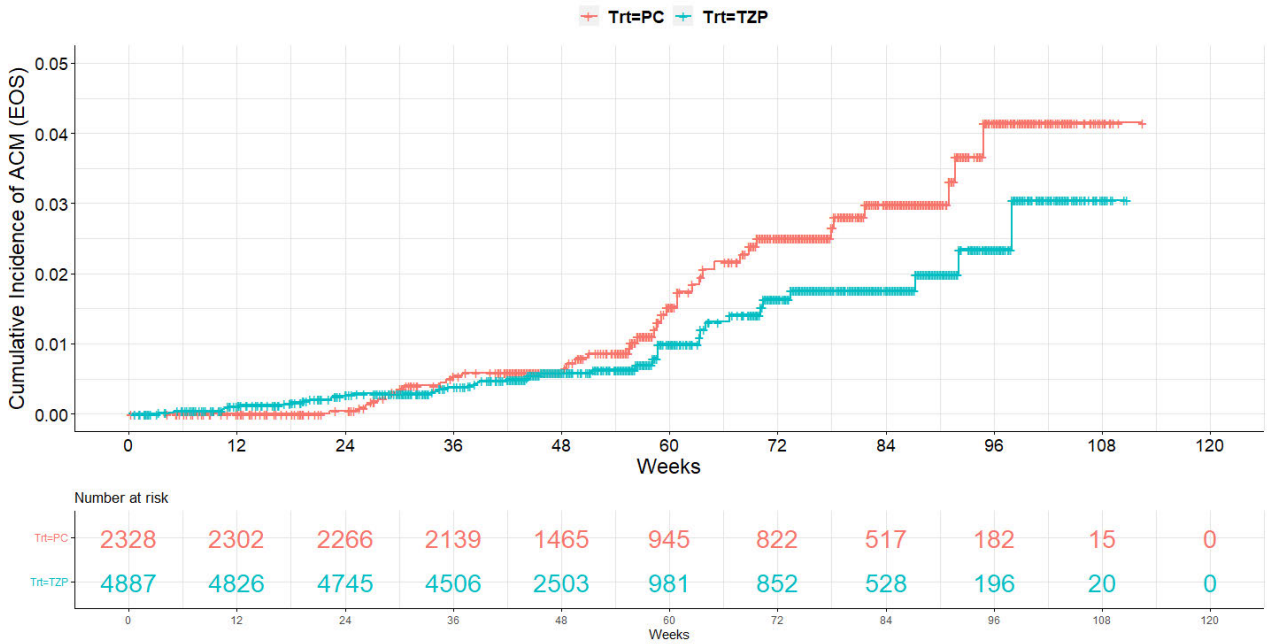
Table 15: Pre-specified Analysis of all-cause mortality Endpoint (mITT, EOMA)*

	Tirzepatide					Pooled Comparator
	15 mg N=1621	**10 mg N=1606**	**5 mg N= 1608**	**1 mg N=52**	**All N= 4887**	**N=2328**
ACM patient years of follow-up	1708.29	1685.82	1676.61	28.67	5099.16	2756.29
ACM events (IR per 100 PY)	13 (0.76)	8 (0.47)	20 (1.19)	0 (0.00)	41 (0.80)	39 (1.41)
HR (95% CI)	0.75 (0.40, 1.41)	0.47 (0.22, 1.01)	1.20 (0.69, 2.08)	0 (0, Inf)	0.80 (0.51, 1.25)	-

* mITT: All randomized subjects who took at least one dose of study drug; EOMA: end-of-meta-analysis.
‡ Hazard ratios for pooled tirzepatide doses computed using a Cox proportional hazards model with fixed effects for treatment and stratified by study-level CV risk (GPGM versus others)

Source: Created by FDA statistical reviewer.

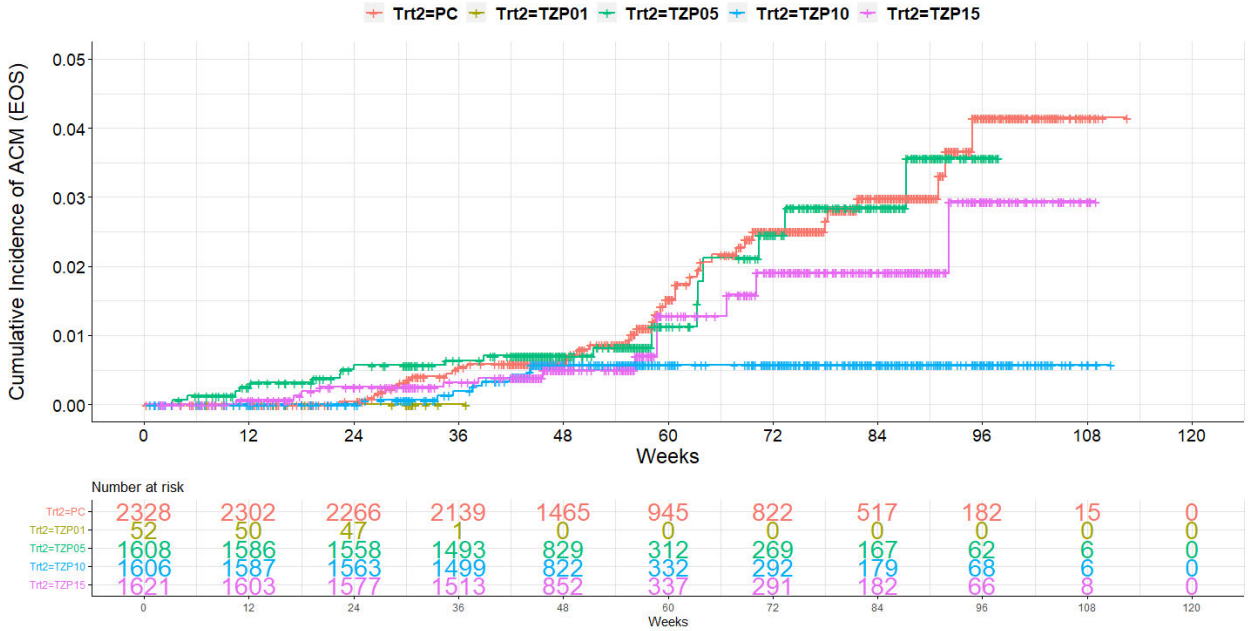
Figure 7: Kaplan-Meier-based cumulative incidence plots for all-cause mortality by pooled treatment arms (mITT, EOMA)*



* mITT: All randomized subjects who took at least one dose of study drug; EOMA: End-of meta-analysis
Source: Created by FDA statistical reviewer.

Reference ID: 4972437

Figure 8: Kaplan-Meier-based cumulative incidence plots for all-cause mortality by treatment arms and tirzepatide dose arms (mITT, EOMA)*



The analyses for all-cause mortality do not raise any concerns of excess mortality in the tirzepatide dose arms over that in the pooled comparator arm.

3.3.4.4 Exploratory Analyses for Pulse Rate Changes

The clinical review team noticed a higher numerical change in pulse rate from baseline to W24 in subjects randomized in Japan, and it requested statistical help to explore possible explanations for this difference. Exploratory analyses of pulse rate changes were conducted for the CV meta-analysis population excluding the Phase 2 study GPGB – i.e., Studies GPGH, GPGI, GPGK, GPGL, GPGM, GPGO were analyzed.

The Week 24 timepoint was recommended by the clinical reviewer as the timepoint of interest – i.e., pulse rate changes from baseline to Week 24 were analyzed.

3.3.4.4.1 Descriptive Analyses of Pulse Rate

Table 16 and Table 17 below presents counts by study and dose for the 6 studies included in these exploratory analyses of pulse rate and counts by Japan[7] and non-Japan subjects.

Table 16: Number of patients by study and dose in meta-analysis Phase 3 studies

	Tirzepatide				Pooled Comparator	Total
Study*	**1 mg**	**5 mg**	**10 mg**	**15 mg**		
I8F-MC-GPGH	0	358	360	359	360	**1437**
I8F-MC-GPGI	0	116	119	120	120	**475**
I8F-MC-GPGK	0	121	121	121	115	**478**
I8F-MC-GPGL	0	470	469	470	469	**1878**
I8F-MC-GPGM	0	329	328	338	1000	**1995**

[7] Patients randomized in Japan who took at least one dose of study treatment.

Reference ID: 4972437

	Tirzepatide				Pooled Comparator	Total
Study*	**1 mg**	**5 mg**	**10 mg**	**15 mg**		
I8F-JE-GPGO	0	159	158	160	159	**636**
Total	**52**	**1608**	**1606**	**1621**	**2328**	**7215**
*Study GPGB was excluded since it is a Phase 2 study.						

Source: Created by the FDA statistical reviewer.

Table 17: Mean and standard deviation of pulse at baseline and Week 24 by trial and Japan/non-Japan subgroups in meta-analysis Phase 3 studies*

	Baseline pulse		Pulse at W24	
Study	**Japan**	**Non-Japan**	**Japan**	**Non-Japan**
GPGH	NA	75.22 (9.69)	NA	77.75 (9.54)
GPGI	76.46 (10.97)	74.9 (11.19)	82.78 (12.47)	77.35 (10.35)
GPGK	74.52 (10.2)	73.54 (9.11)	80.04 (9.72)	75.48 (8.54)
GPGL	NA	74.75 (10.07)	NA	78.6 (9.24)
GPGM	NA	72.76 (10.49)	NA	75.25 (10.42)
GPGO	72.87 (10.34)	NA	78.71 (10.77)	NA
Total	**73.42 (10.44)**	**74.06 (10.2)**	**79.27 (10.91)**	**77.01 (9.86)**
*Study GPGB was excluded since it was a Phase 2 study.				

Source: Created by the FDA statistical reviewer.

3.3.4.4.2 Analyses of Change in Pulse Rate at Week 24

The change in pulse rate at Week 24 was modeled as a function of Dose (linear and quadratic terms), baseline BMI, baseline Pulse, and interactions between Dose and baseline BMI and between Dose and baseline pulse rate. A term for the Japan subset was also fit. Statement of the linear model used (coefficients are omitted) is below.

ChangeAtW24 = Intercept +Dose +Dose^2+BMI+Pulse+BMI x Dose + Pulse x Dose
+Japan + Error

Table 20 presented estimated coefficients and associated p-values for model terms. We see that the Dose, baseline Pulse, Dose x baseline Pulse and Japan coefficients are significant at the unadjusted 5% level. Thus, it would appear that these are potentially strong predictors of pulse rate change at Week 24, i.e., Change in pulse at Week 24 is potentially strongly influenced by Dose, baseline Pulse, the interaction between Dose and baseline Pulse, and being treated in Japan. There is an average decrease of 4.17 units in change in pulse rate at Week 24 for the Japan subset.

Table 18: Summary of Results for model of Week 24 change in pulse rate from baseline for Phase 3 studies included in the meta-analysis including main effect for Japan

Terms	Coeff (p-value)
Intercept	31.93 [<0.001]
Study	O:-1.58[0.04]; H:0.46(0.11); I:0.29 (0.50); K:-0.86 (0.06); L:1.48[<0.001]
Dose†	1.10 [<0.001]
Dose^2	-0.01 (0.08)
BMI.bl	0.01 (0.73)
Pulse.bl	-0.37 [<0.001]
Japan	-4.17 [<0.001]
Dose x BMI.bl	-0.003 (0.35)
Dose x Pulse.bl	-0.009 [<0.001]
[]: model term was significant at the 5% level; (): model term not significant at 5% level (exploratory analyses not	

Reference ID: 4972437

adjusted for multiplicity). Only last letter of studies with significant coefficients are listed in the Study row above.
*GPGM was the reference study for All and non-Japan subgroups; GPGO was reference for Japan subgroup
†Dose values are 0: All non-tirzepatide; 1:tirzepatide 1 mg; 5: tirzepatide 5 mg; 10: tirzepatide 10mg; 15: tirzepatide 15 mg

Source: Created by the FDA statistical reviewer.

Separate models were fit for the overall population, and Japan and non-Japan subsets to see how the models differed. A statement of the linear model used (coefficients are omitted) is below.
ChangeAtW24 = Intercept +Dose +Dose^2+BMI+Pulse+BMI x Dose + Pulse x Dose+Error.

Table 19: Summary of Results for models of Week 24 change in pulse rate from baseline for Phase 3 studies included in the meta-analysis

	Coeff (p-value)		
Terms	**All***	**Japan***	**non-Japan***
Intercept	28.06 [<0.001]	22.99 [<0.001]	29.41 [<0.001]
Study	O:2.53[<0.001], I:1.04 [0.01], L:1.5 [<0.001]	I:1.80 (0.06); K:0.22 (0.80)	H:0.53 (0.07); I:0.28 (0.54); K:-0.74(0.11); L:1.54 p<0.001]
Dose†	1.10 [<0.001]	0.88 (0.09)	1.06 [<0.001]
Dose^2	-0.01 (0.09)	0.01 (0.29)	-0.01[0.02]
BMI.bl	-0.003 (0.90)	0.01 (0.92)	-0.01 (0.70)
Pulse.bl	-0.37 [<0.001]	-0.28 [<0.001]	-0.38 [<0.001]
Dose x BMI.bl	-0.003 (0.29)	-0.01 (0.54)	0.001 (0.80)
Dose x Pulse.bl	-0.009 [<0.001]	-0.01 (0.24)	0.01 [<0.001]

[]: model term was significant at the 5% level; (): model term not significant at 5% level (exploratory analyses not adjusted for multiplicity). Only last letter of studies with significant coefficients are listed in the Study row above.
*GPGM was the reference study for All and non-Japan subgroups; GPGO was reference for Japan subgroup
†Dose values are 0: All non-tirzepatide; 1:tirzepatide 1 mg; 5: tirzepatide 5 mg; 10: tirzepatide 10mg; 15: tirzepatide 15 mg

Source: Created by the FDA statistical reviewer.

Models for the Phase 3 meta-analysis population indicate that Dose and baseline pulse and the interaction of baseline Pulse and Dose are significant factors for change in pulse rate at Week 24 for the full Phase 3 meta-analysis population and non-Japan subgroup. Of these only baseline Pulse is significant for the Japan subgroup. BMI or interactions of BMI with dose are not significant predictors in these models for any of these groups.
The effect of Dose was not consistent in the full meta-analysis population and the Japan subgroup. Baseline Pulse was a significant negative predictor; in the Japan subgroup a unit increase in baseline pulse resulted in an average decrease of 0.28 units in the change in pulse rate at Week 24. The effect of dose was not consistent in the full meta-analysis population and the Japan subset.

Reviewer Comments:
These exploratory results should be interpreted as generating possible hypotheses for future research since these analyses were not pre-specified but were rather data-driven. Hence, we would expect that the type I error is likely higher than that specified by the confidence level. Pre-specification of appropriate hypotheses when blinded to future study data are suggested for confirmatory conclusions.

Reference ID: 4972437

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

Events in the on-treatment + 30-day window in the end-of-study data on the mITT population were used for all the exploratory subgroup analyses in this section. Hazard ratios and associated unadjusted 95% confidence intervals for the subgroups were estimated using Cox proportional hazards (CPH) models with a fixed term for treatment[8] and stratified by individual study. A CPH model, that included a fixed term for treatment, a fixed term for the subgroup of interest and the treatment-subgroup interaction term, with study as stratification factor, was used to check for interaction effects of the subgroup.

4.1 Gender, Race, Age, and Geographic Region Subgroups for MACE-4 endpoint

Table 21 presents the results of subgroup analyses for the MACE-4 endpoint across key baseline demographic characteristics, and Figure 6 contains the corresponding forest plot. Event counts, incidence rates, hazard ratio estimates and associated 95% confidence intervals are presented for subgroups defined by Age, Sex, Race, US/outside US (OUS) categorization and Region using on-study events. Confidence interval for the Age Group >=65 subgroup indicated a potential beneficial effect of pooled tirzepatide arm over pooled comparator arm at the unadjusted 5% level of significance. There were no significant interactions at the unadjusted 5% level of significance for any of the demographic subgroups listed in Table 21.

Table 20: Demographic Subgroup analyses for MACE-4 endpoint (mITT; EOS)

Subgroup	tirzepatide	Placebo	HR† (95% CI)
	No. Events/N (IR per 100 PY)	No. Events/N (IR per 100 PY)	
Age Group (years)			
< 65	45/3421 (1.31)	31/1496 (1.87)	1.06 (0.66, 1.67)
>= 65	27/1466 (1.66)	39/832 (3.68)	0.58 (0.35, 0.95)
Sex			
F	18/2163 (0.82)	18/962 (1.66)	0.65 (0.33, 1.27)
M	54/2724	52/1366	0.87 (0.59, 1.29)
Race			
White	57/3507 (1.55)	53/1783 (2.50)	0.86 (0.59, 1.26)
Black or African American	2/185 (1.17)	3/69 (3.86)	0.51 (0.08, 3.04)
Asian	3/792 (0.37)	2/273 (0.70)	0.71 (0.12, 4.43)
American Indian or Alaskan Native	8/356 (2.29)	8/164 (4.36)	0.65 (0.24, 1.82)
Native Hawaiian or Other Pacific Islander	0/8 (0.00)	1/4 (3.77)	0.00 (0, Inf)
Multiple	1/36 (2.38)	3/32 (6.62)	0.58 (0.06, 5.69)
US/OUS			
US	14/1084 (1.36)	19/503 (3.61)	0.53 (0.26, 1.07)
OUS	58/3803 (1.44)	51/1825 (2.33)	0.90 (0.61, 1.32)
Region			
Asia (excluding Japan)	2/205 (1.00)	2/84 (2.14)	1.02 (0.14, 7.22)

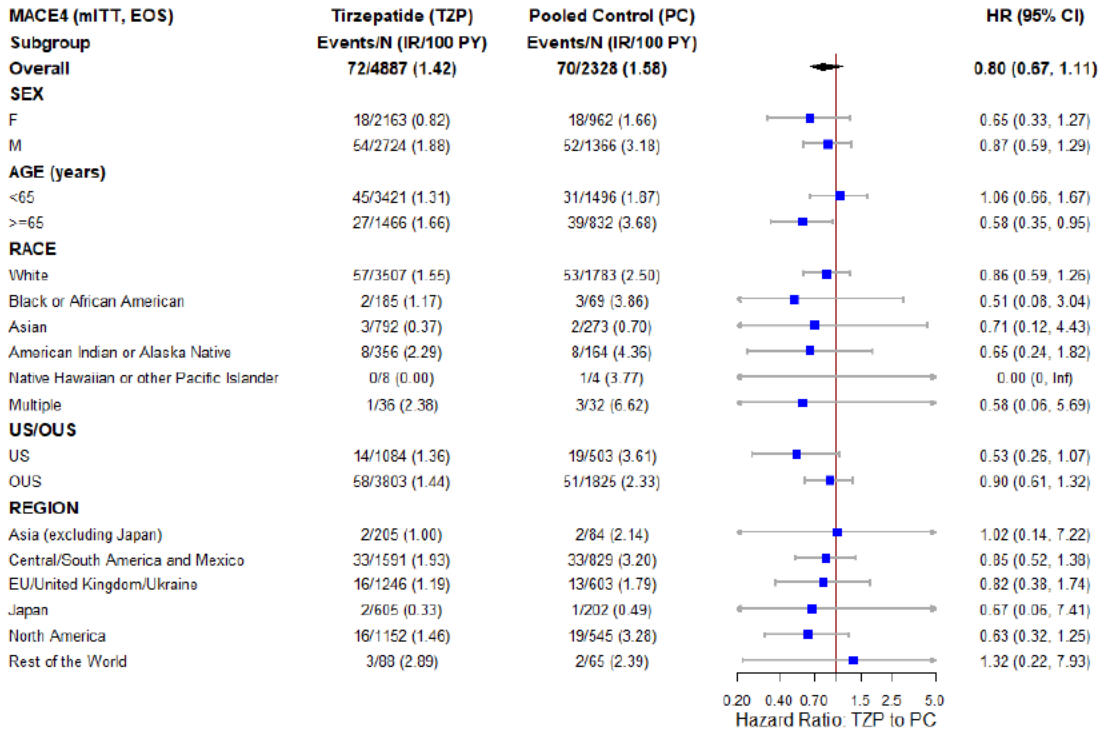
[8] The Cox PH model for the primary MACE4 analysis was stratified by CV risk strata, i.e., Study GPGM patients were in the high-risk stratum and all other subjects were in the low-risk stratum.

Reference ID: 4972437

Subgroup	tirzepatide	Placebo	HR† (95% CI)
	No. Events/N (IR per 100 PY)	No. Events/N (IR per 100 PY)	
Central/South America and Mexico	33/1591 (1.93)	33/829 (3.20)	0.85 (0.52, 1.38)
EU/United Kingdom/Ukraine	16/1246 (1.19)	13/603 (1.79)	0.82 (0.38, 1.74)
Japan	2/605 (0.33)	1/202 (0.49)	0.67 (0.06, 7.41)
North America	16/1152 (1.46)	19/545 (3.28)	0.63 (0.32, 1.25)
Rest of the World	3/88 (2.89)	2/65 (2.39)	1.32 (0.22, 7.93)
†A Cox proportional hazards model with fixed treatment effect was used to estimate hazard ratios (HR) and 95% confidence intervals.			

Source: Created by the FDA statistical reviewer.

Figure 9: Forest plot for demographic Subgroups for MACE-4 endpoint (mITT; EOS)



Source: Created by the FDA statistical reviewer.

4.2 Clinical Subgroups for MACE-4 endpoint

Table 22 presents the results of subgroup analyses for the MACE-4 endpoint across key baseline clinical characteristics, and Figure 7 contains the corresponding forest plot. Event counts, incidence rates, hazard ratio estimates and associated 95% confidence intervals are presented for subgroups based on baseline clinical characteristics. Confidence intervals for some subgroups – Duration of Diabetes Mellitus ≥10 years, eGFR < 60, and UACR indicating microalbuminaria – indicate a potential beneficial effect of pooled tirzepatide over pooled comparator at the unadjusted 5% level of significance. The only significant interaction at the 5% level of significance was that for Duration of DM – the hazard ratio for those with >=10 years of DM

Reference ID: 4972437

was protective at the unadjusted 5% level, while that for patients with Duration of DM <10 years was not. No other clinical subgroups listed in Table 22 had any significant interaction effects.

Table 21: Clinical subgroup analyses for MACE-4 endpoint (mITT; EOS)

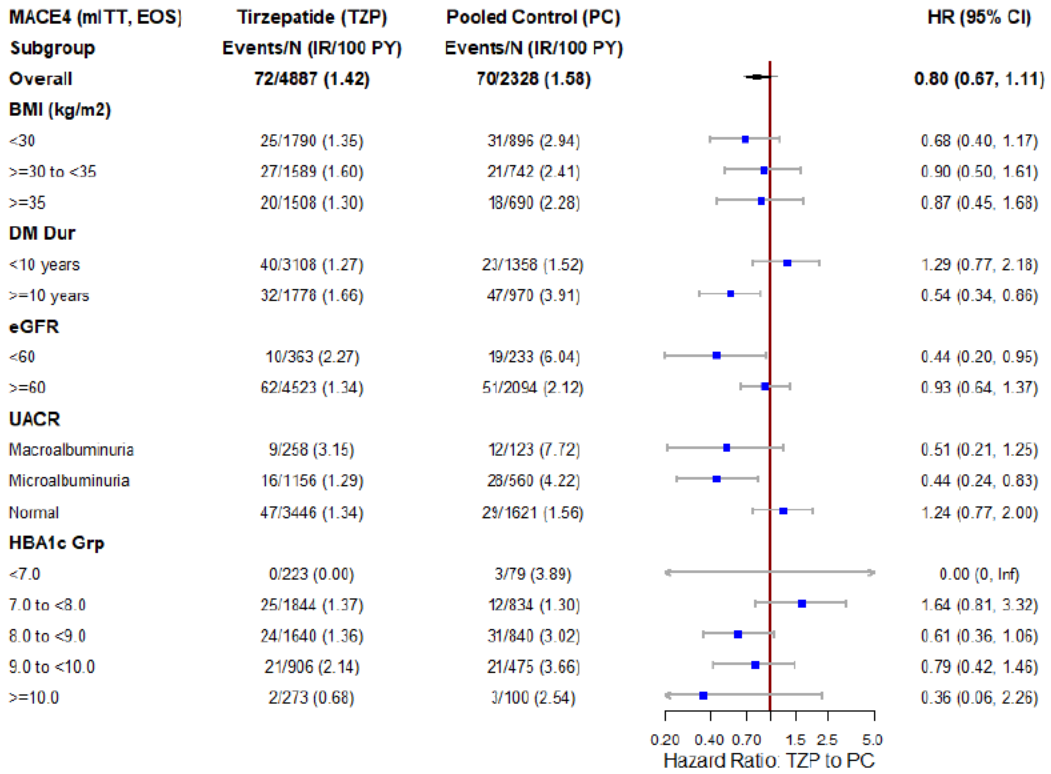
Subgroup	Tirzepatide No. Events/N (IR per 100 PY)	Placebo No. Events/N (IR per 100 PY)	HR[†] (95% CI)
BMI *(kg/m²)*			
< 30	25/1790 (1.35)	31/896 (2.94)	0.68 (0.40, 1.17)
30 to < 35	27/1589 (1.60)	21/742 (2.41)	0.90 (0.50, 1.61)
≥ 35	20/1508 (1.30)	18/690 (2.28)	0.87 (0.45, 1.68)
Duration of DM[♦] *(years)*			
< 10	40/3108 (1.27)	23/1358 (1.52)	1.29 (0.77, 2.18)
≥ 10	32/1778 (1.66)	47/970 (3.91)	0.54 (0.34, 0.86)
eGFR[†]			
<60	10/363 (2.27)	19/233 (6.04)	0.44 (0.20, 0.95)
≥ 60	62/4523 (1.34)	51/2094 (2.12)	0.93 (0.64, 1.37)
UACR			
Macro	9/258 (3.15)	12/123 (7.72)	0.51 (0.21, 1.25)
Micro	16/1156 (1.29)	28/560 (4.22)	0.44 (0.24, 0.83)
Normal	47/3446 (1.34)	29/1621 (1.56)	1.24 (0.77, 2.00)
A1c Group[*]			
< 7	0/223 (0.00)	3/79 (3.89)	0.00 (0, Inf)
7 to <8	25/1844 (1.37)	12/834 (1.30)	1.64 (0.81, 3.32)
8 to <9	24/1640 (1.36)	31/840 (3.02)	0.61 (0.36, 1.06)
9 to < 10	21/906 (2.14)	21/475 (3.66)	0.79 (0.42, 1.46)
≥ 10	2/273 (0.68)	3/100 (2.54)	0.36 (0.06, 2.26)

[†]A Cox proportional hazards model with fixed treatment effect was used to estimated hazard ratios (HR) and 95% confidence intervals.
[§] One patient with missing baseline Tobacco Use was excluded.
[†]Two patients with missing baseline eGFR status were excluded.
[♦] One patient with missing baseline duration of DM was excluded.
[*] One patient with missing baseline A1c Group was excluded.

Source: Created by the FDA statistical reviewer.

Reference ID: 4972437

Figure 10: Forest plot for clinical subgroups for MACE-4 endpoint (mITT; EOS)



Source: Created by the FDA statistical reviewer.

4.3 Gender, Race, Age, and Geographic Region Subgroups for All-Cause Mortality

Table 23 presents the results of subgroup analyses for the all-cause mortality endpoint across key baseline demographic characteristics, and Figure 8 contains the corresponding forest plot. Event counts, incidence rates, hazard ratio estimates and associated 95% confidence intervals are presented for subgroups defined by Age, Sex, Race, US/OUS categorization and Region. Confidence intervals for most subgroups indicate no evidence of either harmful or beneficial effect of pooled tirzepatide over pooled comparator for all-cause mortality at the unadjusted 5% level of significance. The exceptions were: Age Group >=65, and American Indian or Alaska Native Race. There were no significant interactions at the 5% level of significance for any of the demographic subgroups listed in Table 23.

Table 22: Demographic Subgroup analyses for all-cause mortality endpoint (mITT; EOS)

Subgroup	tirzepatide	Placebo	HR† (95% CI)
	No. Events/N (IR per 100 PY)	No. Events/N (IR per 100 PY)	HR† (95% CI)
Age Group (years)			
< 65	24/3421 (0.69)	12/1496 (0.81)	1.40 (0.69, 2.85)
>= 65	17/1466 (1.04)	27/832 (2.51)	0.52 (0.28, 0.97)
Sex			
F	11/2163 (0.5)	12/962 (1.10)	0.58 (0.25, 1.34)

Reference ID: 4972437

Subgroup	tirzepatide	Placebo	HR† (95% CI)
	No. Events/N (IR per 100 PY)	No. Events/N (IR per 100 PY)	HR† (95% CI)
M	30/2724 (1.03)	27/1366 (1.62)	0.93 (0.54, 1.57)
Race§			
White	31/3507 (0.83)	21/1783 (0.98)	1.11 (0.63, 1.95)
Black or African American	2/185 (1.16)	2/69 (2.52)	0.77 (0.11, 5.47)
Asian	0/792 (0.00)	1/273 (0.35)	0.00 (0, Inf)
American Indian or Alaskan Native	7/356 (1.99)	12/164 (6.53)	0.38 (0.14, 1.00)
Native Hawaiian or Other Pacific Islander	0/8 (0.00)	0/4 (4.16)	NA
Multiple	0/36 (0.00)	3/32 (6.61)	0.00 (0, Inf)
US/OUS			
US	6/1084 (0.58)	4/503 (0.75)	0.99 (0.27, 3.60)
OUS	35/3803 (0.86)	35/1825 (1.58)	0.78 (0.48, 1.26)
Region			
Asia (excluding Japan)	0/205 (0.00)	0/84 (0.00)	NA
Central/South America and Mexico	25/1591 (1.45)	31/829 (2.97)	0.62 (0.36, 1.06)
EU/United Kingdom/Ukraine	9/1246 (0.66)	3/603 (0.41)	2.41 (0.64, 9.08)
Japan	0/605 (0.00)	0/202 (0.00)	NA
North America	6/1152 (0.54)	4/545 (0.68)	1.00 (0.27, 3.64)
Rest of the World	1/88 (0.95)	1/65 (1.19)	0.97 (0.06, 15.55)

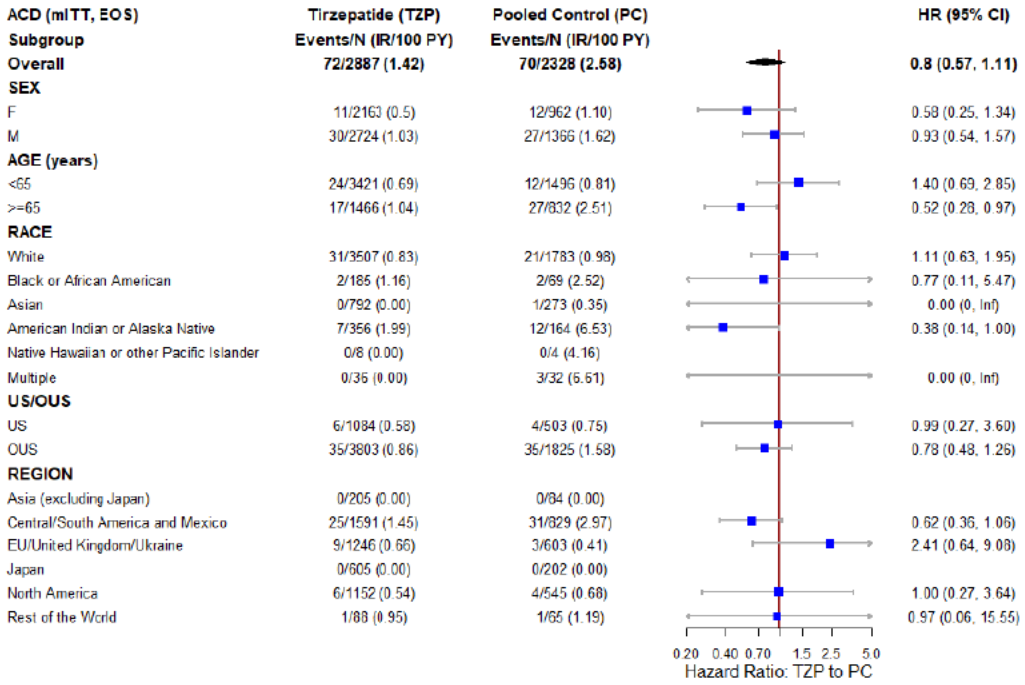
†A Cox proportional hazards model with fixed treatment effect was used to estimate hazard ratios (HR) and 95% confidence intervals.

§ 6 subjects – 3 on tirzepatide and 3 on pooled control arm – who were missing Race values at baseline, were excluded from this analysis.

Source: Created by the FDA statistical reviewer.

Reference ID: 4972437

Figure 11: Forest plot for demographic subgroups for all-cause mortality (mITT; EOS)



Source: Created by the FDA statistical reviewer.

4.4 Clinical Subgroups for All-Cause Mortality

Table 24 presents the results of subgroup analyses for the all-cause mortality (ACM) endpoint across key baseline clinical characteristics, and Figure 9 contains the corresponding forest plot. Event counts, incidence rates, hazard ratio estimates and associated 95% confidence intervals are presented for subgroups based on baseline clinical characteristics. Confidence intervals for none of the subgroups indicated evidence of either harmful or beneficial effect of pooled tirzepatide over pooled comparator arm at the unadjusted 5% level of significance, on all-cause mortality.

There were no significant interactions at the 5% level of significance for any of the subgroups listed in Table 24.

Table 23: Clinical subgroup analyses for ACM endpoint (mITT; EOS)

Subgroup	Tirzepatide No. Events/N (IR per 100 PY)	Placebo No. Events/N (IR per 100 PY)	HR† (95% CI)
BMI *(kg/m²)*			
< 30	15/1790 (0.81)	19/896 (1.78)	0.67 (0.33, 1.34)
30 to < 35	12/1589 (0.71)	13/742 (1.46)	0.69 (0.31, 1.52)
≥ 35	14/1508 (0.91)	7/690 (0.88)	1.36 (0.53, 3.46)
Duration of DM[1] *(years)*			
< 10	23/3108 (0.73)	13/1358 (0.85)	1.27 (0.63, 2.54)
≥ 10	18/1778 (0.93)	26/970 (2.12)	0.55 (0.30, 1.03)

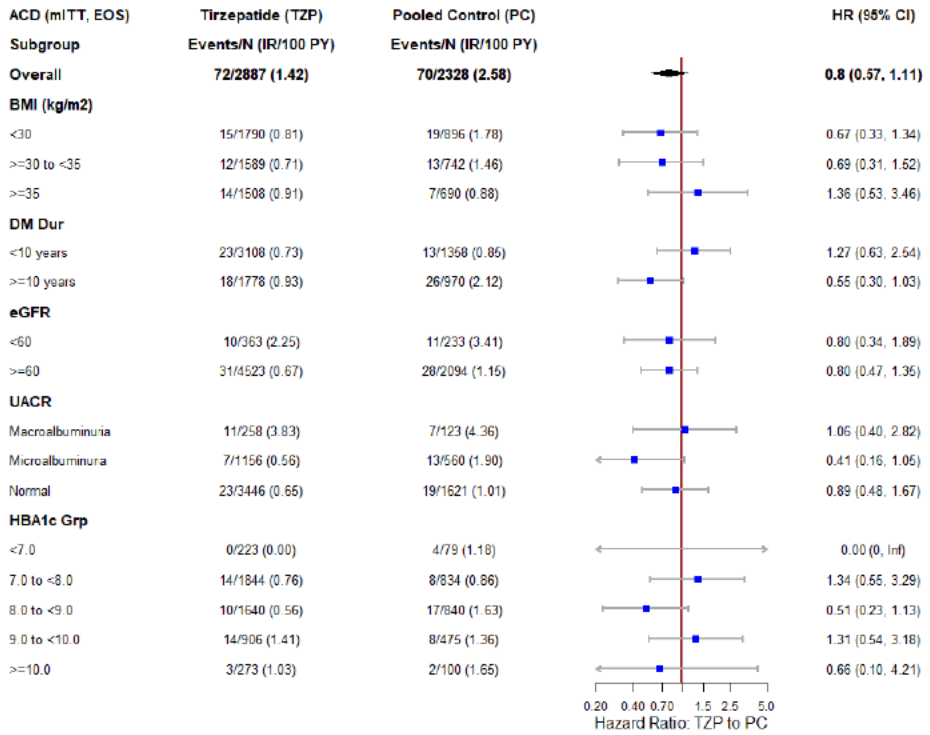
Reference ID: 4972437

Subgroup	Tirzepatide No. Events/N (IR per 100 PY)	Placebo No. Events/N (IR per 100 PY)	HR† (95% CI)
eGFR[2]			
<60	10/363 (2.25)	11/233 (3.41)	0.80 (0.34, 1.89)
≥ 60	31/4523 (0.67)	28/2094 (1.15)	0.80 (0.47, 1.35)
UACR[3]			
Macro	11/258 (3.83)	7/123 (4.36)	1.06 (0.40, 2.82)
Micro	7/1156 (0.56)	13/560 (1.90)	0.41 (0.16, 1.05)
Normal	23/3446 (0.65)	19/1621 (1.01)	0.89 (0.48, 1.67)
A1c Group[4]			
< 7	0/223 (0.00)	4/79 (1.18)	0.00 (0, Inf)
7 to <8	14/1844 (0.76)	8/834 (0.86)	1.34 (0.55, 3.29)
8 to <9	10/1640 (0.56)	17/840 (1.63)	0.51 (0.23, 1.13)
9 to < 10	14/906 (1.41)	8/475 (1.36)	1.31 (0.54, 3.18)
≥ 10	3/273 (1.03)	2/100 (1.65)	0.66 (0.10, 4.21)

†A Cox proportional hazards model with fixed treatment effect was used to estimated hazard ratios (HR) and 95% confidence intervals.
[1]One patient with missing baseline duration of DM was excluded from this analysis.
[2] 2 patients with missing eGFR values at baseline were excluded from this analysis.
[3] 51 patients with missing UACR values at baseline – 27 on the tirzepatide arm and 24 on pooled comparator – were excluded from this analysis.
[4] One patient with missing A1c value at baseline was excluded from this analysis.

Source: Created by the FDA statistical reviewer.

Figure 12: Forest plot for clinical subgroups for all-cause mortality (mITT; EOS)



Source: Created by the FDA statistical reviewer.

Reference ID: 4972437

5 SUMMARY AND CONCLUSIONS

5.1 Collective Evidence

A CV meta-analysis that included 1 Phase 2 and 6 Phase 3 studies was carried out to establish CV safety of tirzepatide. Three of these 7 studies were open label studies.

The pre-specified primary analysis for the primary MACE-4 endpoint was carried out at trial interim when 116 MACE-4 events had been observed and resulted in a hazard ratio estimate of **0.81** and a CI of **(0.52, 1.26)** at the pre-specified 97.85% level. The upper bound of this interval is less than 1.8 and thus meets the pre-specified margin in the meta-analysis statistical analysis plan. Results from supporting analyses on components of the MACE-4 endpoint and the analysis of this endpoint using the complete data were consistent with this conclusion. The active comparators in the different meta-analysis studies, and the fact that 3 of the 7 studies in the meta-analysis were open label studies dictate that these results be interpreted cautiously, as larger double-blind, randomized studies with different, or a different mix of comparators may not result in the same conclusions.

A total of 6579 subjects was analyzed in the interim analysis for the primary safety endpoint, 4-component major adverse cardiovascular event (MACE-4) defined as a composite of the following adjudicated events: cardiovascular death, myocardial infarction, stroke and hospitalization for unstable angina. Study GPGO was not included in this interim look since it had not been completed at the time of database lock for the interim analysis. The pre-specified analysis for the primary MACE-4 endpoint was time until first MACE-4. This analysis resulted in a hazard ratio estimate of 0.81 and a 97.85% confidence interval of (0.52, 1.26).

Analyses of all-cause mortality estimated a hazard ratio of 0.80 with an associated 95% CI of (0.51, 1.25) for the comparison of pooled tirzepatide arms over pooled comparator arm. An analysis of the all-cause mortality endpoint did not raise any concerns of excess mortality in the tirzepatide dose arms over that in the pooled comparator arm.

Due to a clinical concern about pulse rate exploratory analyses were conducted to assess pulse rate changes in the Japan subset at the request of the clinical review team. These analyses indicated differences in the Japan and non-Japan subsets.

5.2 Statistical Issues

The following issues should be taken into account when interpreting the results of the meta-analysis provided to assess CV risk of tirzepatide:

- 3 of the 7 randomized trials in the meta-analysis were open-label trials; knowledge of drug a patient is on could cause bias in results due to differences in patient management and other factors.
- Different comparators – active and placebo – were used in the different studies in the meta-analysis. Study GPGM which contributed most of the MACE-4 events had insulin glargine as the active comparator and showed a numerically beneficial effect for MACE-4 in the tirzepatide arm; Study GPGL which had semaglutide as the active comparator,

Reference ID: 4972437

showed a numerically beneficial effect in the semaglutide arm. Any potential generalization of results from this meta-analysis would have to take this into account.

5.3 Conclusions and Recommendations

The cardiovascular safety of tirzepatide was evaluated based on results of the CV meta-analysis that included 7 trials, 1 Phase 2 trial and 6 Phase 3 trials. Three of the 7 trials were open-label trials whereas the rest were double-blind. A total of 7215 patients was randomized and took at least one dose of study treatment. These subjects contributed a total of 7781.8 patient-years of experience – 5064.45 on the tirzepatide arms and 2717.35 on the pooled comparator arms.

The interim analysis for the primary safety endpoint, 4-component major adverse cardiovascular event (MACE-4) defined as a composite of the following adjudicated events: cardiovascular death, myocardial infarction, stroke and hospitalization for unstable angina, estimated a hazard ratio of 0.81 with an associated 97.85% CI of (0.52, 1.26). The upper bound of this confidence interval is less than 1.8 and meets the risk margin specified by the 2008 FDA Guidance for Industry.

A hazard ratio of 0.80 with an associated 95% CI of (0.51, 1.25) for the endpoint of all-cause mortality was estimated for the comparison of pooled tirzepatide arms over pooled comparator arm. Analyses of the all-cause mortality endpoint did not raise any concerns.

Exploratory analyses were conducted to assess pulse rate changes in the Japan subset at the request of the clinical review team. These analyses of Week 24 change from baseline in pulse rate indicate that Dose, baseline Pulse, and their interaction are significant predictors of Week 24 change in pulse rate. However, there are also differences in Japan and non-Japan subsets – for the Japan subgroup baseline Pulse is the only significant factor among these three. Pre-specification of hypotheses for future studies is recommended for confirmatory conclusions.

Based on this evidence, we consider that the CV meta-analysis was generally successful in demonstrating cardiovascular safety of tirzepatide when compared to the standard of care. The meta-analysis included open label studies and different active comparators; hence these results should be interpreted with caution; they may not be generalizable.

Reference ID: 4972437

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

SHANTI V GOMATAM
04/21/2022 04:08:25 PM

EUGENIO ANDRACA-CARRERA
04/21/2022 04:59:37 PM

MARK S LEVENSON
04/21/2022 05:00:58 PM

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

Drug Name:	Mounjaro (tirzepatide) 5mg, 10mg or 15mg, once weekly injection
Indication(s):	An adjunct to diet and exercise to improve glycemic control in adults with type 2 diabetes mellitus (T2DM)
Applicant:	Elly Lilly
Date(s):	Date Received: 9/16/21
	Filing Date: 11/14/21 (Filing Meeting on 10/18/21)
	Primary Review Goal Date: 2/15/22
	User Fee Goal Date: 5/15/22
Review Priority:	Priority
Biometrics Division:	DB II
Statistical Reviewer:	Wenda Tu
Concurring Reviewers:	Yoonhee Kim (Acting Team Lead)
Medical Division:	DDLO
Clinical Team:	Frank Pucino
Project Manager:	Lindsey Kelly

Keywords:

ANCOVA, Washout Imputation, Imputation based on Retrieved Dropouts, Shrinkage Analyses, Bayesian Hierarchical Modelling, NME, Dual GIP and GLP-1

Table of Contents

1 EXECUTIVE SUMMARY 5

1.1 BRIEF OVERVIEW OF CLINICAL STUDY 5
1.2 MAJOR STATISTICAL ISSUES 5
1.3 COLLECTIVE EVIDENCE 5
1.4 CONCLUSION AND RECOMMENDATIONS 6

2 INTRODUCTION 7

2.1 OVERVIEW 7
2.2 DATA SOURCES 8

3 STATISTICAL EVALUATION 8

3.1 DATA AND ANALYSIS QUALITY 8
3.2 EVALUATION OF EFFICACY 8
3.2.1 Study Design and Endpoints 8
3.2.2 Statistical Methodologies 19
3.2.3 Patient Disposition, Demographic and Baseline Characteristics 20
3.2.4 Results and Conclusions 28
3.3 EVALUATION OF SAFETY 32

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS 35

4.1 GENDER, RACE, AGE, AND GEOGRAPHIC REGION 35
4.2 EFFECT MODIFICATION OF BASELINE HBA1C 41

5 SUMMARY AND CONCLUSIONS 42

5.1 STATISTICAL ISSUES 42
5.2 COLLECTIVE EVIDENCE 42
5.3 CONCLUSIONS AND RECOMMENDATIONS 43
5.4 LABELING RECOMMENDATIONS 43

Reference ID: 4938321

LIST OF TABLES

Table 1: Analysis Results for HbA1c (%) Change from Baseline at Week 40 (SUPRASS-1, -2, and -5) or Week 52 (SURPASS-3 and -4) 6
Table 2: Patient Disposition, All Randomized Population, SURPASS-1 21
Table 3: Patient Disposition, All Randomized Population, SURPASS-2 21
Table 4: Patient Disposition, All Randomized Population, SURPASS-3 22
Table 5: Patient Disposition, All Randomized Population, SURPASS-3 22
Table 6: Patient Disposition, All Randomized Population, SURPASS-5 23
Table 7: Patient Demographics and Baseline Characteristics, All Randomized Patients, SURPASS-1 24
Table 8: Patient Demographics and Baseline Characteristics, All Randomized Patients, SURPASS-2 25
Table 9: Patient Demographics and Baseline Characteristics, All Randomized Patients, SURPASS-3 26
Table 10: Patient Demographics and Baseline Characteristics, All Randomized Patients, SURPASS-4 27
Table 11: Patient Demographics and Baseline Characteristics, All Randomized Patients, SURPASS-5 28
Table 12: Analysis Results for HbA1c (%) Change from Baseline at Week 40 29
Table 13: Sensitivity Analyses of the Primary Endpoint Based on the return-to-baseline Method 30
Table 14: Analysis Results for Weight (kg) Change from Baseline 30
Table 15: Analysis Results for FSG (mg/dL) Change from Baseline 31
Table 16: Analysis Results for Incidence of HbA1c < 7% 31
Table 17: Summary and Analysis of Level 2 or Level 3 Hypoglycemia Incidence and Rate (SURPASS-2, Modified Intent-to-Treat Population, Safety Analysis Set) 33
Table 18: Summary and Analysis of Level 2 or Level 3 Hypoglycemia Incidence and Rate (SURPASS-3, Modified Intent-to-Treat Population, Safety Analysis Set) 33
Table 19: Summary and Analysis of Level 2 or Level 3 Hypoglycemia Incidence and Rate (SURPASS-4, Modified Intent-to-Treat Population, Safety Analysis Set) 34
Table 20: Summary and Analysis of Level 2 or Level 3 Hypoglycemia Incidence and Rate (SURPASS-5, Modified Intent-to-Treat Population, Safety Analysis Set) 34
Table 21: Sample and shrinkage Estimates of Difference in HbA1c % Change from Baseline within Subgroups, mITT Population, SURPASS-1 36
Table 22: Sample and Shrinkage Estimates of Difference in HbA1c % Change from Baseline within Subgroups, mITT Population, SURPASS-2 36
Table 23: Sample and shrinkage Estimates of Difference in HbA1c % Change from Baseline within Subgroups, mITT Population, SURPASS-3 37
Table 24: Sample and Shrinkage Estimates of Difference in HbA1c % Change from Baseline within Subgroups, mITT Population, SURPASS-4 37
Table 25: Sample and shrinkage Estimates of HbA1c % Change from Baseline within Subgroups, mITT Population, SURPASS-5 38
Table 26: Slopes (S.E.) of Linear Regressions (A1c change from baseline ~ baseline A1c) Base on Completers Dataset 41

Reference ID: 4938321

LIST OF FIGURES

Figure 1: Dose Escalation Scheme for the Phase 3 Program....8
Figure 2: Graphical Testing Hierarchy for SURPASS-1....10
Figure 3: Graphical Testing Hierarchy for SURPASS-2....12
Figure 4: Graphical Testing Hierarchy for SURPASS-3....14
Figure 5: Graphical Testing Hierarchy for SURPASS-4....16
Figure 6: Graphical Testing Hierarchy for SURPASS-5....18
Figure 7: Subgroup Analysis for Age (< 65 Years, >= 65 Years), SURPASS-2....38
Figure 8: Subgroup Analysis for Race, SURPASS-2....39
Figure 9: Subgroup Analysis for Sex, SURPASS-2....39
Figure 10: Subgroup Analysis for Region (US or Outside US), SURPASS-2....40
Figure 11: A1c Change from Baseline (%) vs Baseline A1c (%), Completers Data, SURPASS-1....41
Figure 12: A1c Change from Baseline (%) vs Baseline A1c (%), Completers Data, SURPASS-5....42

Reference ID: 4938321

1 EXECUTIVE SUMMARY

The applicant, Eli Lilly and Company, submitted this original new drug application (NDA) for tirzepatide (TZP), a new molecular entity with the proposed indication for adults with type 2 diabetes mellitus (T2DM). The product was studied under three doses: 5mg, 10mg and 15 mg, and was applied either as a monotherapy, in combination with metformin, sulfonylureas (SU), and SGLT-2 inhibitors (SGLT-2i) alone/combined, or in combination with basal insulin with or without metformin.

1.1 Brief overview of Clinical Study

This submission included five Phase III trials: SURPASS-1 (GPGK), SURPASS-2 (GPGL), SURPASS-3 (GPGH), SURPASS-4 (GPGM), and SURPASS-5 (GPGI). For each study, subjects were randomized to one of the four arms: TZP 5mg, TZP 10mg, TZP 15mg or the comparator arm, at a randomization ratio of 1:1:1:1 (except for SURPASS-4, which was 1:1:1:3). SURPASS-1 and -5 were the two placebo-controlled trials: SURPASS-1 assessed the drug as a monotherapy, whereas SURPASS-5 evaluated the product as an add-on to insulin glargine with or without metformin. SURPASS-2 compared the product with semaglutide 1 mg, where both treatments were applied as add-ons to metformin. SURPASS-3 compared the product with insulin degludec, both as add on to metformin with or without SGLT-2i. SRUPASS-4 compared the product with insulin glargine, in which one, two, or three background therapies from metformin, SU or SGLT-2i were applied to the enrolled subjects. The treatment period for SURPASS-3 and -4 were of 52 weeks, while the treatment period for other three studies were of 40 weeks. The primary efficacy endpoint for all five trials was HbA1c change from baseline. More information about the study designs can be found in Section 2.1.

1.2 Major Statistical Issues

No major statistical issues have been identified. For efficacy evaluation, the applicant applied the treatment-policy estimand with missing endpoint measurements imputed based on data collected either from retrieved dropouts, or from placebo arms if insufficient retrieved dropouts (Section 3.2.2). Sensitivity analyses based on return-to-baseline imputation have been performed for all five studies (Section 3.2.2).

As a minor issue, for subgroup analyses, significant interactions with *age* and *race* were detected in some of the studies. However, these interactions were due to differences in effect sizes instead of differences in effect directions. Shrinkage analyses were performed to double check these interaction effects (Section 4.1).

1.3 Collective Evidence

Results from the primary efficacy analyses demonstrated statistically significant superiority of tirzepatide to placebo, semaglutide 1 mg, insulin degludec and insulin glargine. Table 1 provided the key findings from primary analyses (see Table 12 for more details). Results from the sensitivity analyses exhibited consistent efficacy results as the primary analyses (Table 13).

Reference ID: 4938321

Analyses on the key secondary endpoints (including weight change from baseline, fasting serum glucose change from baseline, and incidence of HbA1c < 7%) also displayed favorable outcomes, and hence provided additional evidence to support the findings on the primary endpoint (Tables 14, 15 and 16). Subgroup analyses on the primary efficacy endpoint suggested that the efficacy of tirzepatide were not impacted by age (< 65 years or ≥ 65 years), gender (Male or Female), race (Asian, Black, or White, etc.) and region (the US or outside the US) (Tables 21 through 25). In addition, analyses on the safety database did not find an increased risk of Level 2 or Level 3 hypoglycemic events among the tirzepatide-treated subjects compared to subjects treated with placebo, semaglutide, insulin degludec, or insulin glargine (Tables 17 through 20).

Table 1: Analysis Results for HbA1c (%) Change from Baseline at Week 40 (SUPRASS-1, -2, and -5) or Week 52 (SURPASS-3 and -4)

	SURPASS-1	**SURPASS-2**	**SURPASS-3**	**SURPASS-4**	**SURPASS-5**
Comparator	**Placebo**	**Semaglutide 1 mg**	**Insulin degludec**	**Insulin glargine**	**Placebo**
Comparator-adjusted treatment effect with 95% CI					
TZP 5mg	-1.64*** (-1.94, -1.34)	-0.16* (-0.29, -0.04)	-0.54*** (-0.78, -0.30)	-0.76*** (-0.89, -0.64)	-1.25*** (-1.49, -1.01)
TZP 10mg	-1.60*** (-1.90, -1.30)	-0.41*** (-0.53, -0.28)	-0.73*** (-0.98, -0.48)	-0.92*** (-1.04, -0.79)	-1.53*** (-1.77, -1.30)
TZP 15mg	-1.59*** (-1.90, -1.28)	-0.48*** (-0.60, -0.36)	-0.85*** (-1.10, -0.60)	-1.02*** (-1.14, -0.89)	-1.48*** (-1.72, -1.24)

P-values (two-sided) for superiority: *p-value < 0.05, *** p-value < 0.001 vs placebo or active comparator.

1.4 Conclusion and Recommendations

Statistical analyses based on the clinical data collected from the five phase III trials SURPASS-1 through -5 have demonstrated robust evidence in support of efficacy of the tirzepatide in treating adults with T2DM. In particular, statistical findings have shown compelling effect size and robust superiority of tirzepatide 5mg, 10mg and 15mg with respect to glycemic control, when compared to placebo, semaglutide 1mg, insulin glargine and insulin degludec. This statistical reviewer recommends an approval of the proposed indication of tirzepatide 5 mg, 10mg, and 15 mg for adults with T2DM.

Reference ID: 4938321

2 INTRODUCTION

2.1 Overview

Tirzepatide (TZP) is a dual gastric inhibitory polypeptide (GIP) and glucagon-like peptide-1 receptor agonist (GLP-1 RA), a new drug class with intended indication for glycemic control for adults with T2DM. This original submission contains five pivotal phase III trials designed to evaluate the efficacy, safety, and tolerability of once-weekly treatment with injectable tirzepatide at maintenance doses of 5mg, 10mg and 15mg among adult subjects with T2DM. In these studies, tirzepatide was compared against either placebo or active comparators, and was assessed as monotherapy or add-on treatment to oral antidiabetic medicine or basal insulin. Key factors of the five studies were summarized in Table 1.

Table 1: Clinical Studies Reviewed by the Statistical Reviewer

Trial ID	**GPGK SURPASS-1**	**GPGL SURPASS-2**	**GPGH SURPASS-3**	**GPGM* SURPASS-4**	**GPGI SURPASS-5**
Design	Double blind	Open label	Open label	Open label	Double blind
Comparator	Placebo	Semaglutide	Insulin degludec	Insulin glargine	Placebo
Background Medications	None (lifestyle changes only)	Metformin	Metformin ± SGLT-2i	1 to 3 OAMs (± metformin ± SU ± SGLT-2i)	Insulin glargine ± metformin
Randomization Scheme	1:1:1:1	1:1:1:1	1:1:1:1	1:1:1:3	1:1:1:1
# of Randomized and Treated patients[1]	475	1876	1435	1989	471
Treatment Period Duration	40 weeks	40 weeks	52 weeks	52 weeks + (52--104 weeks for long-term safety period)	40 weeks
Endpoints	***Primary*:** change from baseline in A1c at week 40 ***Key Secondary:*** Incidence of A1c < 7% Change from baseline in body weight Incidence of A1c < 5.7% Change from baseline in FSG	***Primary:*** change from baseline in A1c at week 40 ***Key Secondary:*** Incidence of A1c < 7% Change from baseline in body weight Incidence of A1c < 5.7%	***Primary:*** change from baseline in A1c at week 52 ***Key Secondary:*** Incidence of A1c < 7% Change from baseline in body weight	***Primary*:** change from baseline in A1c at week 52 ***Key Secondary:*** Incidence of A1c < 7% Change from baseline in body weight	***Primary*:** change from baseline in A1c at week 40 ***Key Secondary:*** Incidence of A1c < 7% Change from baseline in body weight Incidence of A1c < 5.7% Change from baseline in FSG

OAM: Oral Antidiabetic Medication; FSG Fasting Serum Glucose

[1] Excluding inadvertently enrolled participants, which were defined as randomized participants who did not meet the inclusion criteria or met an exclusion criterion

Reference ID: 4938321

2.2 Data Sources

The Electronic Document Room (EDR) locations for the original submission is \\CDSESUB1\evsprod\NDA215866\0001. The datasets (both in ADAM format and SDTM format) and the programming codes for the primary and key secondary efficacy analyses can be found under the subdirectory: m5\datasets.

Upon the Agency's request, sensitivity analyses (both the programming codes and the analysis results) were submitted under SDN 21, with EDR location: \\CDSESUB1\evsprod\NDA215866\0017.

Upon the Agency's request, the programming codes for subgroup analyses were submitted under SDN 25, with EDR location: \\CDSESUB1\evsprod\NDA215866\0028.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

No issues have been identified with respect to data and analysis quality.

3.2 Evaluation of Efficacy

Across the five studies, the dose-escalation scheme was identical. For participants randomized to the tirzepatide arm, the dose-escalation scheme was illustrated as below.

Treatment Group	Treatment Period Intervals					
	Weeks 0 to 4	Weeks 4 to 8	Weeks 8 to 12	Weeks 12 to 16	Weeks 16 to 20	Week 20 through End of Treatment Period
Tirzepatide 5 mg	2.5 mg	5 mg	5 mg	5 mg	5 mg	5 mg
Tirzepatide 10 mg	2.5 mg	5 mg	7.5 mg	10 mg	10 mg	10 mg
Tirzepatide 15 mg	2.5 mg	5 mg	7.5 mg	10 mg	12.5 mg	15 mg

Figure 1: Dose Escalation Scheme for the Phase 3 Program

For each trial, study design, primary/key secondary endpoints, and multiple testing scheme with hypotheses are described in the following subsections.

3.2.1 Study Design and Endpoints

SURPASS-1 (GPGK)

The trial was a multi-center, multinational, randomized, double-blind, parallel-group, placebo-controlled study designed to assess the efficacy and safety of three doses of once weekly

Reference ID: 4938321

tirzepatide (5, 10, 15 mg) compared with placebo in subjects with T2DM who had inadequate glycemic control with diet and exercise alone, had not been treated with any oral antihyperglycemic medication during the three months prior to the start of the study, and were naive to antihyperglycemic injectable therapy. The trial consisted of a 3-week Screening/Lead in period, a 40-week Treatment Period and a 4-week Follow-up Period. Eligible participants were randomized 1:1:1:1 to once weekly injectable tirzepatide 5mg, 10mg, 15mg or placebo. The primary objective was to demonstrate superiority of once weekly TZP 5mg to placebo, TZP 10mg to placebo, and TZP15 mg to placebo on HbA1c change from baseline at 40 weeks.

Sample Size

The trial was powered to assess superiority of TZP 5mg, 10mg, or 15mg vs placebo relative to mean change from baseline in HbA1c at 40 weeks under the following assumptions:

- TZP 5mg, 10mg and 15 mg will be tested in parallel, each at a 2-sided significance level of 0.0017 (*i.e.*, $\alpha/3$ with $\alpha=0.05$)
- At least 0.65% difference in mean reduction in HbA1c between a TZP arm and the placebo arm,
- A common standard deviation of 1.3%.

Based on these assumptions, 472 patients at randomization ratio 1:1:1:1 will provide at least 90% power to establish superiority for a tirzepatide dose compared to placebo.

In the study, a total of 478 subjects were randomized: 121 to the TZP 5mg, 121 to the TZP 10mg, 121 to the TZP 15mg, and 115 to the placebo arm. It appeared that the study had adequate power to detect superiority of the treatment effect of tirzepatide (see details in Section 3.2.4).

Primary Endpoint

- Change from baseline in HbA1c (%)

Key Secondary Endpoints

- Incidence of HbA1c < 7%
- Incidence of HbA1c < 5.7%
- Change from baseline in fasting serum glucose (FSG)
- Change from baseline in body weight (kg)

Multiplicity adjustment

To control a two-sided family-wise type I error of 5%, sequentially rejective graphical procedures[2] were used for multiple study objectives.

The primary and key secondary objective hypotheses for SURPASS-1 were as follows,

- H5,1, H10,1, and H15,1: Superiority test of tirzepatide 5 mg, 10 mg, 15 mg versus placebo in HbA1c change from baseline at 40 weeks respectively.
- H5,2, H10,2, and H15,2: Superiority test of tirzepatide 5 mg, 10 mg, 15 mg versus placebo in body weight change from baseline at 40 weeks respectively.

[2] Xi, D., & Bretz, F. (2021). Graphical Approaches for Multiple Comparison Procedures. In Handbook of Multiple Comparisons (pp. 91-119). CRC Press.

Reference ID: 4938321

- H5,3, H10,3, and H15,3: Superiority test of tirzepatide 5 mg, 10 mg, 15 mg versus placebo in proportion of patients achieve HbA1c<7% at 40 weeks respectively.
- H5,4, H10,4, and H15,4: Superiority test of tirzepatide 5 mg, 10 mg, 15 mg versus placebo in fasting serum glucose (FSG) change from baseline at 40 weeks respectively.
- H5,5, H10,5, and H15,5: Superiority test of tirzepatide 5 mg, 10 mg, 15 mg versus placebo in proportion of patients achieve HbA1c<5.7% at 40 weeks respectively.

H5,1, H10,1, and H15,1 were initially tested each at 0.01667 significance level. The graphical testing scheme as presented in Figure 2 was used to strongly control for type 1 error.

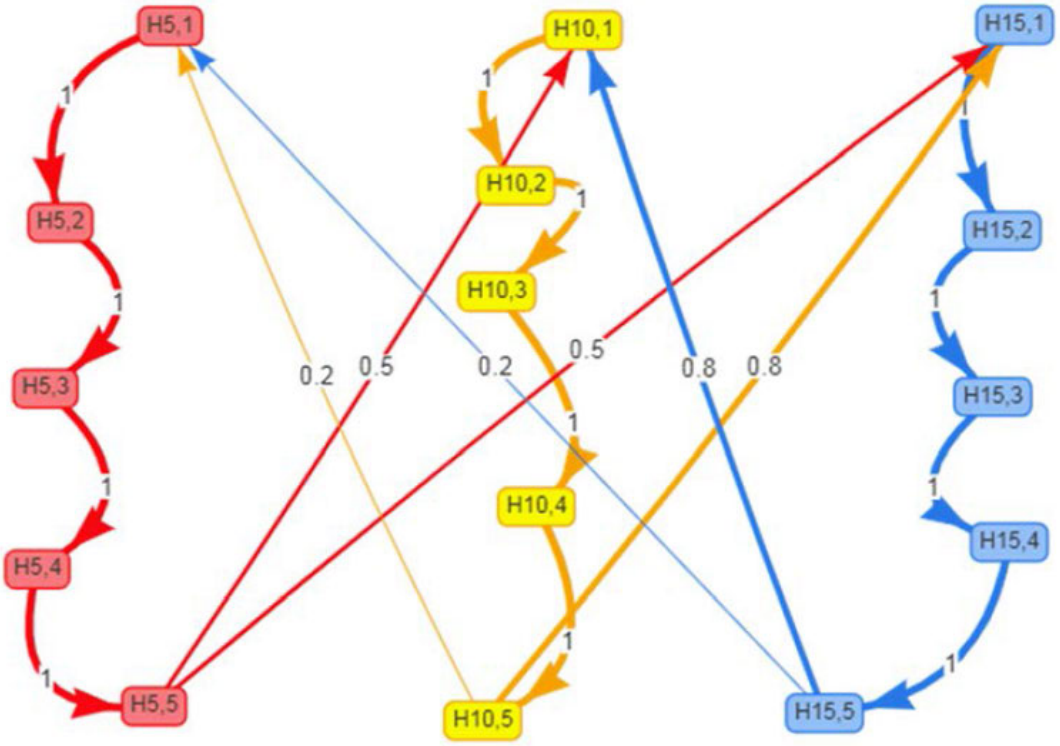


Figure 2: Graphical Testing Hierarchy for SURPASS-1

SURPASS-2 (GPGL)

The trial was a multi-center, multinational, randomized, open-label, parallel-group, active-controlled study designed to assess the efficacy and safety of three once-weekly doses of tirzepatide (5, 10, 15 mg) compared with once-weekly, subcutaneous semaglutide (1mg) in patients with type 2 diabetes who had inadequate glycemic control with metformin monotherapy (≥1500 mg/day) and had not been treated with any other oral antihyperglycemic medications (OAMs) during the three months prior to the start of the study. The trial consisted of a 3-week Screening/Lead in period, a 40-week Treatment Period and a 4-week Follow-up Period. Eligible

Reference ID: 4938321

participants were randomized 1:1:1:1 to once weekly injectable tirzepatide 5mg, 10mg, 15mg or semaglutide 1mg. The primary objective was to demonstrate non-inferiority of tirzepatide 10mg and/or15mg to semaglutide on HbA1c change from baseline at 40 weeks.

Sample Size

The trial was powered to assess non-inferiority of TZP 5mg and/or10mg to semaglutide relative to mean change from baseline in HbA1c at 40 weeks under the following assumptions:

- TZP 10mg and 15mg will be tested in parallel against semaglutide, each at a 2-sided 0.025 significance level (*i.e.*, $\alpha/2$, $\alpha=0.05$)
- No difference in mean reduction in HbA1c between the TZP arm and the comparator arm
- An NI margin of 0.3%
- A common standard deviation of 1.3%

Based on these assumptions, 1872 patients at randomization ratio 1:1:1:1 will provide at least 90% power to demonstrate NI of tirzepatide compared to semaglutide.

In the study, a total of 1879 subjects were randomized: 471 to the TZP 5mg, 469 to the TZP 10mg, 470 to the TZP 15mg, and 469 to the semaglutide arm. It appeared that the study had adequate power to detect superiority of the treatment effect of tirzepatide (see details in Section 3.2.4).

Primary Endpoint

- Change from baseline in HbA1c (%)

Key Secondary Endpoints

- Change from baseline in HbA1c (%) (for superiority tests on 10mg and/or 15mg TZP and non-inferiority & superiority tests on 5mg TZP)
- Incidence of HbA1c < 7%
- Change from baseline in body weight (kg)
- Incidence of HbA1c < 5.7%

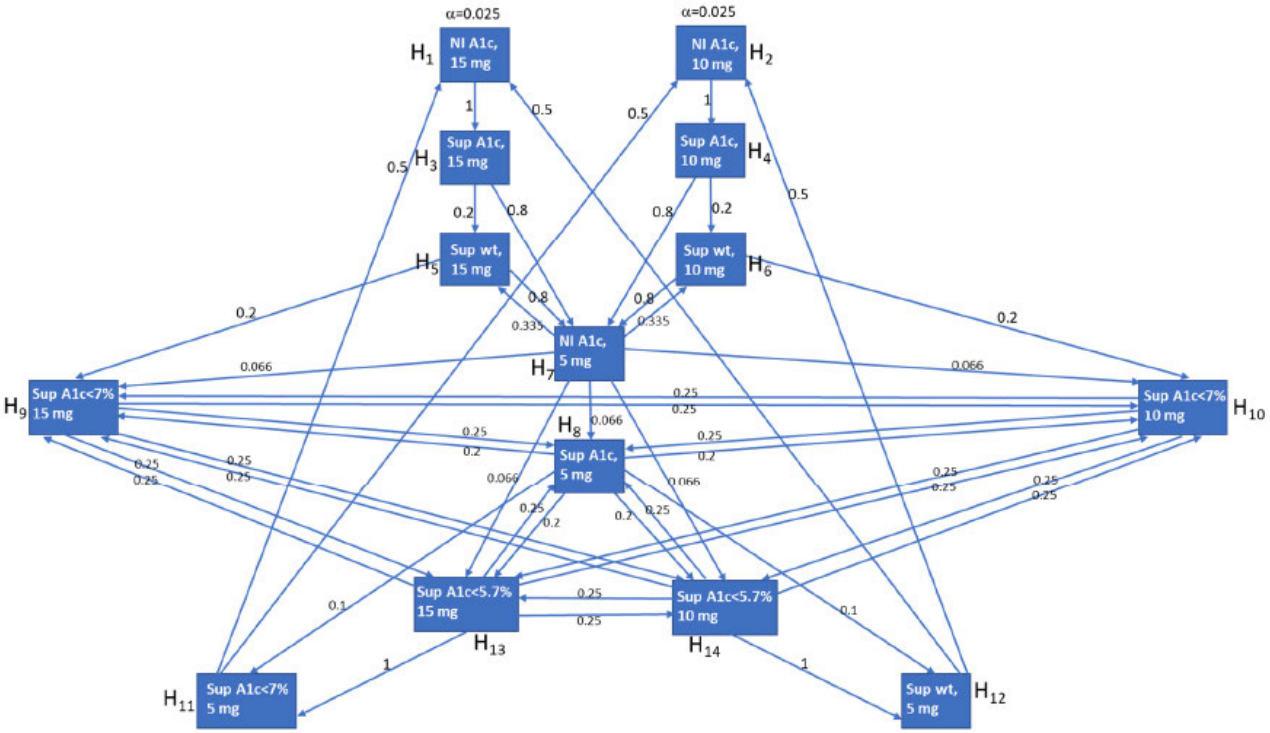
Multiplicity Adjustment

The primary and key secondary objective hypotheses, and the graphical testing scheme were presented as follows.

- H1 and H2: Noninferiority test of tirzepatide 10 mg and 15 mg versus semaglutide in HbA1c change from baseline at 40 weeks.
- H3 and H4: Superiority test of tirzepatide 10 mg, and 15 mg versus semaglutide in HbA1c change from baseline at 40 weeks.
- H5 and H6: Superiority test of tirzepatide 10 mg, and 15 mg versus semaglutide in body weight change from baseline at 40 weeks.
- H7: Noninferiority test of tirzepatide 5 mg versus semaglutide in HbA1c change from baseline at 40 weeks.
- H8: Superiority test of tirzepatide 5 mg versus semaglutide in HbA1c change from baseline at 40 weeks.
- H9, H10, and H11: Superiority test of tirzepatide 10 mg, 15 mg, and 5 mg versus semaglutide in proportion of patients achieving HbA1c <7% at 40 weeks.

Reference ID: 4938321

- H12: Superiority test of tirzepatide 5 mg versus semaglutide in body weight change from baseline at 40 weeks.
- H13 and H14: Superiority test of tirzepatide 10 mg, and 15 mg versus semaglutide in proportion of patients achieving HbA1c < 5.7% at 40 weeks.



Abbreviations: A1c =HbA1C; NI =noninferiority; Sup = superiority; wt =body weight.

Figure 3: Graphical Testing Hierarchy for SURPASS-2

SURPASS-3 (GPGH)

The trial was a multi-center, multinational, randomized, open-label, parallel-group, active-controlled study designed to assess the efficacy and safety of three once-weekly doses of tirzepatide (5, 10, 15 mg) compared with titrated insulin degludec in patients with type 2 diabetes naïve of insulin treatment who had inadequate glycemic control on stable doses of metformin with or without a sodium-glucose cotransporter-2 inhibitor (SGLT-2i). The trial consisted of a 3-week Screening/Lead in period, a 52-week Treatment Period and a 4-week Follow-up Period. During the Treatment Period, eligible subjects were randomized 1:1:1:1 to once weekly injectable tirzepatide 5mg, 10mg, 15mg or once-daily injectable insulin degludec. The primary objective was to demonstrate non-inferiority of tirzepatide 10mg and/or15mg to insulin degludec on HbA1c change from baseline at 40 weeks.

<u>Sample Size</u>

Reference ID: 4938321

The trial was powered to assess superiority of TZP 10mg and/or 15mg to insulin degludec relative to mean change from baseline in HbA1c at 40 weeks under the following assumptions:

- TZP 10mg and 15mg will be tested in parallel against insulin degludec, each at a 2-sided 0.025 significance level (i.e., $\alpha/2$, $\alpha=0.05$)
- A 0.35% difference in mean reduction in HbA1c between a TZP arm and the comparator arm[3]
- A common standard deviation of 1.3%

Based on these assumptions, 1420 patients at randomization ratio 1:1:1:1 will provide at least 90% power to demonstrate superiority of tirzepatide 10mg and/or 15mg compared to insulin degludec.

In the study, a total of 1444 subjects were randomized: 359 to the TZP 5mg, 361 to the TZP 10mg, 359 to the TZP 15mg, and 365 to the insulin degludec arm. It appeared that the study had adequate power to detect superiority of the treatment effect of tirzepatide (see details in Section 3.2.4).

<u>Primary Endpoint</u>

- Change from baseline in HbA1c (%)

<u>Key Secondary Endpoints</u>

- Change from baseline in HbA1c (%) (for non-inferiority tests on 5mg TZP and superiority tests on 5mg, 10mg and/or 15mg TZP)
- Change from baseline in body weight (kg)
- Incidence of HbA1c < 7%

<u>Multiplicity Adjustment</u>

The graphical testing scheme for the primary and the key secondary objectives was presented as below:

[3] The non-inferiority margin was pre-specified as 0.3%.

Reference ID: 4938321

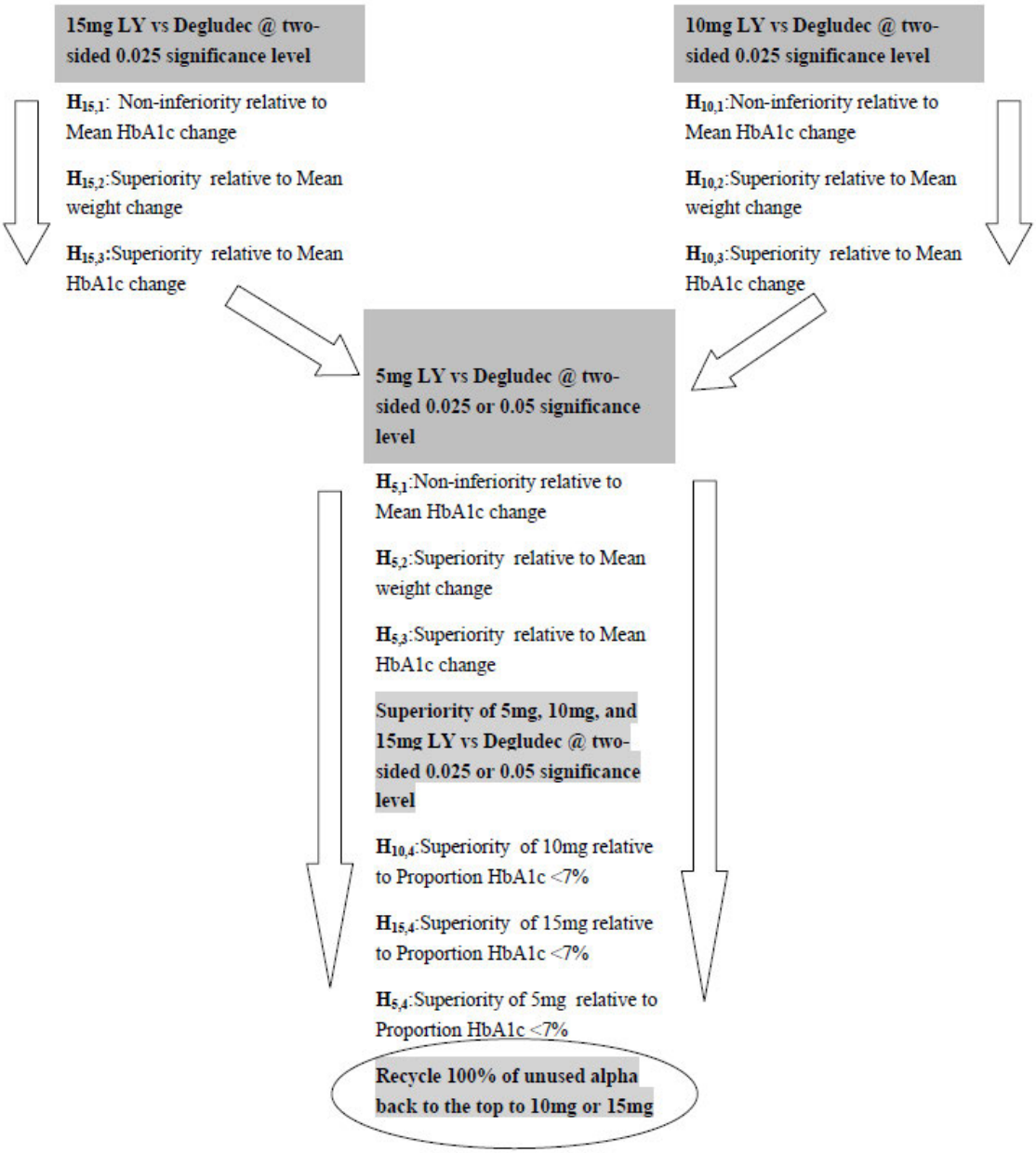


Figure 4: Graphical Testing Hierarchy for SURPASS-3

Reference ID: 4938321

SURPASS-4 (GPGM)

The trial was a multi-center, multinational, randomized, open-label, parallel-group, active-controlled study designed to assess the efficacy and safety of three once-weekly doses of tirzepatide (5, 10, 15 mg) compared with insulin glargine in patients with type 2 diabetes with increased cardiovascular (CV) risk who had inadequate glycemic control on stable doses of 1, 2 or 3 oral antihyperglycemic drugs, including metformin, SGLT-2i, and/or sulfonylurea. The trial consisted of a 2-week Screening/Lead in period, a 52-week Treatment Period I, a Treatment Period II of variable durations (starting at 52 weeks up to 104 weeks), and a 4-week Follow-up Period. Eligible subjects were randomized 1:1:1:3 to once weekly injectable tirzepatide 5mg, 10mg, 15mg or titrated insulin glargine. The primary objective was to demonstrate non-inferiority of tirzepatide 10mg and/or15mg to insulin glargine on HbA1c change from baseline at 52 weeks.

Sample Size

The trial was powered to establish superiority of TZP 10mg and/or 15mg to insulin glargine relative to mean change from baseline in HbA1c at 52 weeks under the following assumptions:

- TZP 10mg and 15mg will be tested in parallel against insulin glargine, each at a 2-sided 0.025 significance level (i.e., $\alpha/2$, $\alpha=0.05$)
- A 0.30% difference in mean reduction in HbA1c between a TZP arm and the comparator arm[4]
- A common standard deviation of 1.3%

Based on these assumptions, 1878 patients randomized to TZP 5mg, TZP 10mg, TZP 15mg, and insulin glargine at randomization ratio 1:1:1:3 will provide at least 90% power to demonstrate superiority of tirzepatide compared to insulin glargine.

In the study, a total of 2002 subjects were randomized: 329 to the TZP 5mg, 330 to the TZP 10mg, 338 to the TZP 15mg, and 1005 to the insulin glargine arm. It appeared that the study had adequate power to detect superiority of the treatment effect of tirzepatide (see details in Section 3.2.4).

Primary Endpoint

- Change from baseline in HbA1c (%)

Key Secondary Endpoints

- Change from baseline in HbA1c (%) (for non-inferiority tests on 5mg TZP and superiority tests on 5mg, 10mg and/or 15mg TZP)
- Change from baseline in body weight (kg)
- Incidence of HbA1c < 7%

Multiplicity Adjustment

The graphical testing scheme for the primary and the key secondary objectives was presented as below:

[4] The non-inferiority margin was pre-specified as 0.3%.

Reference ID: 4938321

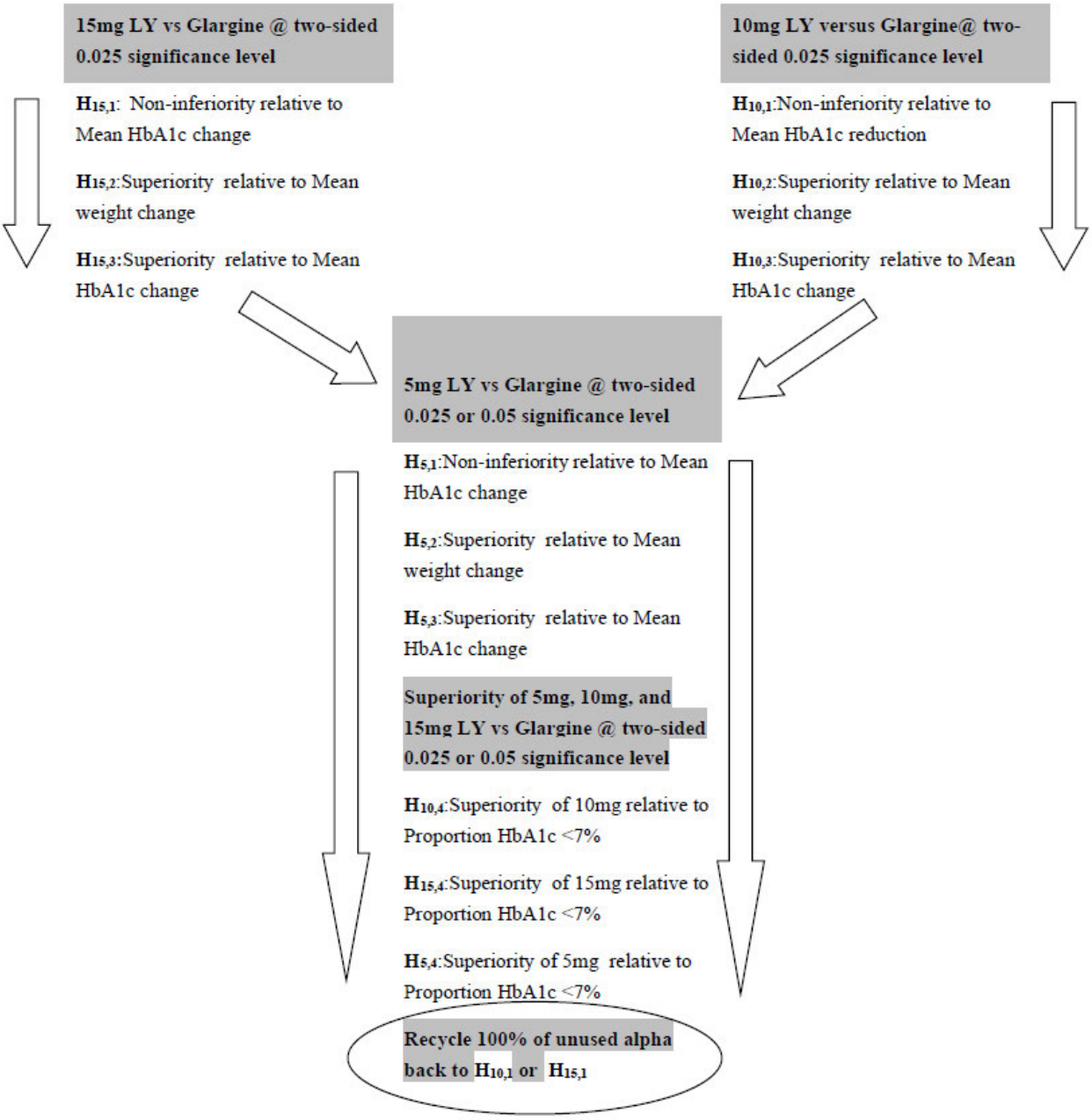


Figure 5: Graphical Testing Hierarchy for SURPASS-4

Reference ID: 4938321

SURPASS-5 (GPGI)

The trial was a multi-center, multinational, randomized, double-blind, parallel-group, placebo-controlled study designed to assess the efficacy and safety of three once-weekly doses of tirzepatide (5, 10, 15 mg) compared with placebo in patients with type 2 diabetes, as an add-on to titrated basal insulin glargine with or without metformin. The trial consisted of a 3-week Screening/Lead in period, a 40-week Treatment Period and a 4-week Follow-up Period. Eligible participants were randomized 1:1:1:1 to once weekly injectable tirzepatide 5mg, 10mg, 15mg or placebo. The primary objective was to demonstrate superiority of once weekly tirzepatide 5mg, 10mg, and/or 15 mg to placebo on HbA1c change from baseline at 40 weeks.

Sample Size

The trial was powered to establish superiority of 10mg and/or 15mg to placebo relative to mean change from baseline in HbA1c at 40 weeks under the following assumptions:

- TZP 10mg and 15mg will be tested in parallel against placebo, each at a 2-sided 0.025 significance level (i.e., $\alpha/2$, $\alpha=0.05$)
- A 0.60% difference in mean reduction in HbA1c between a TZP arm and the comparator arm
- A common standard deviation of 1.3%

Based on these assumptions, 472 patients at randomization ratio 1:1:1:1 will provide at least 90% power to demonstrate superiority of tirzepatide compared to placebo.

In the study, a total of 475 subjects were randomized: 116 to the TZP 5mg, 119 to the TZP 10mg, 120 to the TZP 15mg, and 120 to the placebo arm. It appears that the study had adequate power to detect superiority of the treatment effect of tirzepatide (see details in Section 3.2.4).

Primary Endpoint

- Change from baseline in HbA1c (%)

Key Secondary Endpoints

- Incidence of HbA1c < 7%
- Change from baseline in fasting serum glucose (FSG)
- Change from baseline in body weight (kg)

Multiplicity Adjustment

The graphical testing scheme for the primary and the key secondary objectives was presented as below:

Reference ID: 4938321

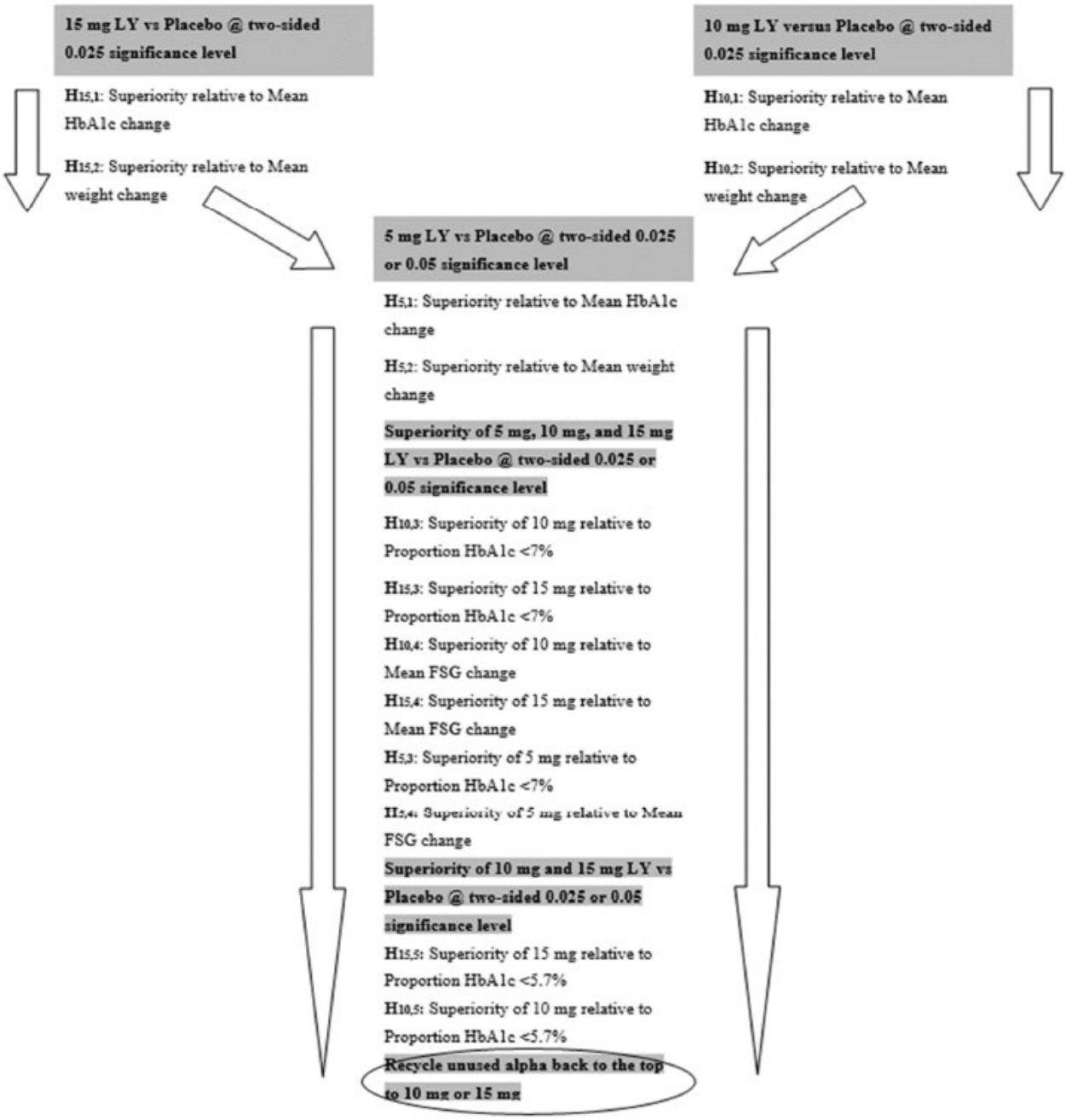


Figure 6: Graphical Testing Hierarchy for SURPASS-5

Reference ID: 4938321

3.2.2 Statistical Methodologies

Population & Analysis Set

For both the primary and key secondary analyses, the target population was the modified intention-to-treat (mITT) population, defined as all randomized subjects who took at least one dose of the study drug, excluding inadvertently enrolled participants[5] as per agreement with the Agency. The analysis set was the full analysis set (FAS), defined as all available data obtained during the Treatment Period (or Treatment Period I for SURPASS-4), regardless of adherence to study drug or initiation of rescue medication.

Handling of Missing Data

For missing primary endpoint measures, multiple imputation based on retrieved dropouts was used in SURPASS-2, -3 and -4. Specifically, for each arm within a study, a regression model adjusted for baseline HbA1c measurement was constructed based on observed data from subjects in the same arm who discontinued treatments but still had their endpoints measured. Missing data were imputed as random draws from a normal distribution centered at the value predicted by the regression model, and with variance set to the variance of the predicted value. The imputation was repeated 100 times to generate 100 complete datasets. Finally, the ANCOVA for the primary efficacy endpoint (as specified in the next section) was performed for each complete dataset, and Rubin's Rule was applied for combing the results for inference.

The placebo-based multiple imputation (or "placebo imputation" as stated by the applicant in the clinical study report (CSR)) was used for SURPASS-1 and -5 due to insufficient retrieved dropout data. Particularly, for each study, a regression model adjusted for baseline HbA1c was built based on completers (*i.e.*, subjects with non-missing primary endpoint measures) from the placebo arm. Next, the same multiple imputation and inference procedure as specified in the retrieved dropout method was performed.

In addition to the applicant's imputation, the reviewer also performed independent multiple imputation based on slightly different imputation models. For SURPASS-2, -3 and -4, the reviewer's model (also based on retrieved dropouts) adjusted for both baseline A1c values and intermediate A1c values. For SURPASS-1 and -5, the reviewer performed multiple imputation based on the washout method; specifically, for each study, the treatment arm model adjusted for baseline HbA1c and all the other covariates specified in the corresponding primary efficacy analysis, whereas the placebo arm model adjusted for baseline A1c, the covariates specified in the primary efficacy analysis, and the intermediate HbA1c values.

Primary & Key Secondary Efficacy Analyses

For each study, an ANCOVA model adjusted for treatment, baseline HbA1c (%), country/pooled country, and past/baseline use of antidiabetic medication[6] was used for the primary efficacy analysis. Analyses of *continuous* key secondary endpoints used ANCOVAs adjusted for

[5] inadvertently enrolled participants were defined as randomized participants who failed the inclusion criteria or met an exclusion criterion.

[6] The SURPASS-1 model adjusted for past use of OAMs (Y/N). The SURPASS-3 and -4 models adjusted for baseline SGLT-2i use (Y/N). The SURPASS -5 model adjusted for baseline metformin use (Y/N).

Reference ID: 4938321

treatment, baseline measures of the variable, country/pooled country, and baseline A1c category[7], whereas analyses of *dichotomous* key secondary endpoints used logistic regressions adjusted for the same set of covariates as the corresponding primary efficacy analyses.

Sensitivity Analysis

Sensitivity Analyses with missing data imputed based on the return-to-baseline method were performed to assess the robustness of the primary analysis results. Specifically, for each study, missing primary endpoint measures from both treatment and control arms were imputed as random draws from a normal distribution centered at zero (the assumed change from baseline value), and with variance set to the residual variance of the ANCOVA used for the primary efficacy analysis but based on completers' data only.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

Summaries of patient disposition of the five study are presented in Tables 2 through 6. For most studies (except for SURPASS-3), patients treated with TZP 15mg showed a higher percentage of *treatment* discontinuation compared to patients who received other treatments. The most likely reasons for treatment discontinuation from TZP 15mg were "withdrawal by subjects" in SURPASS-1 and SURPASS-5, and "adverse events" in SURPASS-2, -3, and -4. No consistent pattern regarding *study* discontinuation was observed across the five studies.

The overall missing rate of the primary endpoints ranged from 4.2% to 9.6% across the five studies (with missing rate ranging from 4% to 8.8% for TZP 5mg, from 2.5% to 10% for TZP 10mg, 4.4% to 14% for TZP 15mg, and 1.7% to 11.3% for the comparator). In SURPASS-2, -3 and -4, there were sufficient retrieved dropout data for missing data imputation. For SURPASS-1 and -5, the retrieved dropout data was limited for some treatment arms, and the placebo-based method (as specified in Section 3.2.2) was used for missing data imputation instead.

[7] Baseline A1c category: ≤ 8.5% or > 8.5% for SURPASS-1, -2, -3, and -4, ≤ 8.0% or > 8.0% for SURPASS-5

Reference ID: 4938321

Table 2: Patient Disposition, All Randomized Population, SURPASS-1

SURPASS-1 (GPGK)	TZP 5mg (N = 121)	TZP 10mg (N = 121)	TZP 15mg (N = 121)	Placebo (N = 115)	Total (N = 478)
Randomized and treated	**121 (100.0)**	**121 (100.0)**	**121 (100.0)**	**115 (100.0)**	**478 (100.0)**
Inadvertent enrollment	0 (0.0)	0 (0.0)	1 (0.8)	2 (1.7)	3 (0.6)
Completed study	114 (94.2)	112 (92.6)	103 (85.1)	99 (86.1)	428 (89.5)
Completed treatment	110 (90.9)	109 (90.1)	95 (78.5)	98 (85.2)	412 (86.2)
Without rescue medication	108 (89.3)	105 (86.8)	92 (76.0)	70 (60.9)	375 (78.5)
With rescue medication	2 (1.7)	4 (3.3)	3 (2.5)	28 (24.3)	37 (7.7)
Discontinued study prior to the primary endpoint visit	**6 (5.0)**	**7 (5.8)**	**15 (12.4)**	**12 (10.4)**	**40 (8.4)**
Adverse event	1 (0.8)	3 (2.5)	1 (0.8)	1 (0.9)	6 (1.3)
Lost to follow-up	2 (1.7)	1 (0.8)	2 (1.7)	4 (3.5)	9 (1.9)
Physician decision	0 (0.0)	1 (0.8)	0 (0.0)	2 (1.7)	3 (0.6)
Protocol deviation	1 (0.8)	0 (0.0)	0 (0.0)	1 (0.9)	2 (0.4)
Withdrawal by subject	2 (1.7)	2 (1.7)	12 (9.9)	4 (3.5)	20 (4.2)
Discontinued treatment	**11 (9.1)**	**12 (9.9)**	**26 (21.5)**	**17 (14.8)**	**66 (13.8)**
Adverse event	4 (3.3)	6 (5.0)	8 (6.6)	2 (1.7)	20 (4.2)
Death	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.9)	1 (0.2)
Failure to meet inclusion criteria	0 (0.0)	0 (0.0)	1 (0.8)	0 (0.0)	1 (0.2)
Lost to follow-up	3 (2.5)	1 (0.8)	2 (1.7)	4 (3.5)	10 (2.1)
Physician decision	0 (0.0)	2 (1.7)	3 (2.5)	4 (3.5)	9 (1.9)
Protocol deviation	1 (0.8)	0 (0.0)	0 (0.0)	1 (0.9)	2 (0.4)
Withdrawal by subject	2 (1.7)	3 (2.5)	11 (9.1)	5 (4.3)	21 (4.4)
Other	1 (0.8)	0 (0.0)	1 (0.8)	0 (0.0)	2 (0.4)
Retrieved dropouts*	**4 (3.3)**	**3 (2.3)**	**8 (6.6)**	**2 (1.7)**	**17 (3.6)**
Missed primary endpoints*	**7 (5.8)**	**9 (7.4)**	**17 (14.0)**	**13 (11.3)**	**46 (9.6)**

* Information on "Retrieved Dropouts" and "Missed Primary Endpoints" were provided by the reviewer based on the datasets: adsl and ada1c. Patients inadvertently enrolled in the study were not counted in either category.

Source: Table GPGK.8.2, CSR & reviewer's analysis.

Table 3: Patient Disposition, All Randomized Population, SURPASS-2

SURPASS-2 (GPGL)	TZP 5mg (N = 471)	TZP 10mg (N = 469)	TZP 15mg (N = 470)	Semaglutide (N = 469)	Total (N = 1879)
Randomized and treated	**470 (99.8)**	**469 (100.0)**	**470 (100.0)**	**469 (100)**	**1878 (99.9)**
Inadvertent enrollment	0 (0.0)	0 (0.0)	1 (0.2)	1 (0.2)	2 (0.1)
Completed study	452 (96.0)	442 (94.2)	446 (94.9)	443 (94.5)	1783 (94.9)
Completed treatment	431 (91.5)	411 (87.6)	408 (86.8)	428 (91.3)	1678 (89.3)
Without rescue medication	424 (90.0)	405 (86.4)	402 (85.5)	416 (88.7)	1647 (87.7)
With rescue medication	7 (1.5)	6 (1.3)	6 (1.3)	12 (2.6)	31 (1.6)
Discontinued study prior to the primary endpoint visit	**14 (3.0)**	**16 (3.4)**	**17 (3.6)**	**18 (3.8)**	**65 (3.5)**
Adverse event	1 (0.2)	4 (0.9)	1 (0.2)	2 (0.4)	8 (0.4)
Death	4 (0.8)	2 (0.4)	2 (0.4)	1 (0.2)	9 (0.5)
Lost to follow-up	3 (0.6)	3 (0.6)	7 (1.5)	7 (1.5)	20 (1.1)
Physician decision	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.4)	2 (0.1)
Pregnancy	1 (0.2)	0 (0.0)	1 (0.2)	1 (0.2)	3 (0.2)
Protocol deviation	0 (0.0)	1 (0.2)	0 (0.0)	0 (0.0)	1 (0.1)
Termination by Sponsor	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)	1 (0.1)
Withdrawal by subject	5 (1.1)	4 (0.9)	5 (1.1)	4 (0.9)	18 (1.0)
Other	0 (0.0)	2 (0.4)	1 (0.2)	0 (0.0)	3 (0.2)
Discontinued treatment	**39 (8.3)**	**58 (12.4)**	**62 (13.2)**	**41 (8.7)**	**200 (10.6)**
Adverse event	24 (5.1)	36 (7.7)	37 (7.9)	18 (3.8)	115 (6.1)
Death	4 (0.8)	4 (0.9)	3 (0.6)	1 (0.2)	12 (0.6)
Failure to meet inclusion criteria	0 (0.0)	0 (0.0)	1 (0.2)	1 (0.2)	2 (0.1)

Reference ID: 4938321

Lost to follow-up	4 (0.8)	4 (0.9)	8 (1.7)	9 (1.9)	25 (1.3)
Physician decision	0 (0.0)	3 (0.6)	0 (0.0)	2 (0.4)	5 (0.3)
Pregnancy	1 (0.2)	0 (0.0)	1 (0.2)	1 (0.2)	3 (0.2)
Protocol deviation	0 (0.0)	1 (0.2)	0 (0.0)	1 (0.2)	2 (0.1)
Withdrawal by subject	6 (1.3)	7 (1.5)	7 (1.5)	7 (1.5)	27 (1.4)
Other	0 (0.0)	3 (0.6)	5 (1.1)	1 (0.2)	9 (0.5)
Retrieved dropouts*	**20 (4.2)**	**34 (7.2)**	**39 (8.3)**	**15 (3.2)**	**108 (5.7)**
Missed primary endpoints*	19 (4.0)	24 (5.1)	22 (4.7)	25 (5.3)	90 (4.8)

* Information on "Retrieved Dropouts" and "Missed Primary Endpoints" were provided by the reviewer based on the datasets: adsl and ada1c. Patients inadvertently enrolled in the study were not counted in either category.

Source: Table GPGL.8.3, CSR & reviewer's analysis

Table 4: Patient Disposition, All Randomized Population, SURPASS-3

SURPASS-3 (GPGH)	**TZP 5mg (N = 359)**	**TZP 5mg (N = 361)**	**TZP 5mg (N = 359)**	**Degludec (N = 365)**	**Total (N = 1444)**
Randomized and treated	**358 (99.7)**	**360 (99.7)**	**359 (100.0)**	**360 (98.6)**	**1437 (99.5)**
Inadvertent enrollment	0 (0.0)	0 (0.0)	1 (0.3)	1 (0.3)	2 (0.1)
Completed study	333 (92.8)	321 (88.9)	340 (94.7)	331 (90.7)	1325 (91.8)
Completed treatment	316 (88.0)	294 (81.4)	300 (83.6)	320 (87.7)	1230 (85.2)
Without rescue medication	312 (86.9)	291 (80.6)	294 (81.9)	319 (87.4)	1216 (84.2)
With rescue medication	4 (1.1)	3 (0.8)	6 (1.7)	1 (0.3)	14 (1.0)
Discontinued study prior to the primary endpoint visit	**20 (5.6)**	**29 (8.0)**	**15 (4.2)**	**31 (8.5)**	**95 (6.6)**
Adverse event	5 (1.4)	7 (1.9)	3 (0.8)	1 (0.3)	16 (1.1)
Death	1 (0.3)	2 (0.6)	1 (0.3)	1 (0.3)	5 (0.3)
Lost to follow-up	6 (1.7)	4 (1.1)	4 (1.1)	5 (1.4)	19 (1.3)
Other	0 (0.0)	1 (0.3)	0 (0.0)	3 (0.8)	4 (0.3)
Physician decision	0 (0.0)	1 (0.3)	2 (0.6)	1 (0.3)	4 (0.3)
Protocol deviation	1 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
Withdrawal by subject	7 (1.9)	14 (3.9)	5 (1.4)	20 (5.5)	46 (3.2)
Discontinued treatment	**42 (11.7)**	**66 (18.3)**	**59 (16.4)**	**40 (11.0)**	**207 (14.3)**
Adverse event	26 (7.2)	35 (9.7)	39 (10.9)	5 (1.4)	105 (7.3)
Death	0 (0.0)	2 (0.6)	0 (0.0)	0 (0.0)	2 (0.1)
Failure to meet inclusion criteria	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.3)	1 (0.1)
Lost to follow-up	6 (1.7)	5 (1.4)	4 (1.1)	4 (1.1)	19 (1.3)
Other	1 (0.3)	2 (0.6)	1 (0.3)	2 (0.5)	6 (0.4)
Physician decision	0 (0.0)	2 (0.6)	5 (1.4)	2 (0.5)	9 (0.6)
Protocol deviation	1 (0.3)	0 (0.0)	1 (0.3)	0 (0.0)	2 (0.1)
Withdrawal by subject	8 (2.2)	20 (5.5)	9 (2.5)	26 (7.1)	63 (4.4)
Retrieved dropouts*	**19 (5.3)**	**30 (8.3)**	**39 (10.9)**	**8 (2.2)**	**96 (6.7)**
Missed primary endpoints*	**23 (6.4)**	**36 (10.0)**	**19 (5.3)**	**31 (8.5)**	**109 (7.5)**

* Information on "Retrieved Dropouts" and "Missed Primary Endpoints" were provided by the reviewer based on the datasets: adsl and ada1c. Patients inadvertently enrolled in the study were not counted in either category.

Source: Table GPGH.8.3, CSR & reviewer's analysis.

Table 5: Patient Disposition, All Randomized Population, SURPASS-3

SURPASS-4 (GPGM)	**TZP 5mg (N = 329)**	**TZP 5mg (N = 330)**	**TZP 5mg (N = 338)**	**Glargine (N = 1005)**	**Total (N = 2002)**
Randomized and treated	**329 (100.0)**	**328 (99.4)**	**338 (100.0)**	**1000 (99.5)**	**1995 (99.7)**
Inadvertent enrollment	1 (0.3)	2 (0.6)	1 (0.3)	2 (0.2)	6 (0.3)
Completed study	294 (89.4)	312 (94.5)	313 (92.6)	882 (87.8)	1801 (90.0)
Completed treatment	278 (84.5)	284 (86.1)	283 (83.7)	861 (85.2)	1706 (85.2)
Without rescue medication	277 (84.2)	283 (85.8)	281 (83.1)	856 (85.2)	1697 (84.8)
With rescue medication	1 (0.3)	1 (0.3)	2 (0.6)	5 (0.5)	9 (0.4)

Reference ID: 4938321

Discontinued study prior to the primary endpoint visit	**21 (6.4)**	**11 (3.3)**	**9 (2.7)**	**52 (5.2)**	**93 (4.6)**
Adverse event	3 (0.9)	2 (0.6)	0 (0.0)	3 (0.3)	8 (0.4)
Screen failure	0 (0.0)	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.0)
Death	6 (1.8)	2 (0.6)	2 (0.6)	9 (0.9)	19 (0.9)
Lost to follow-up	4 (1.2)	2 (0.6)	2 (0.6)	11 (1.1)	19 (0.9)
Other	2 (0.6)	2 (0.6)	0 (0.0)	2 (0.2)	6 (0.3)
Physician decision	1 (0.3)	1 (0.3)	0 (0.0)	3 (0.3)	5 (0.2)
Protocol deviation	0 (0.0)	1 (0.3)	0 (0.0)	1 (0.1)	2 (0.1)
Withdrawal by subject	5 (1.5)	0 (0.0)	5 (1.5)	23 (2.3)	33 (1.6)
Discontinued treatment	**51 (15.5)**	**44 (13.3)**	**55 (16.3)**	**139 (13.8)**	**289 (14.4)**
Adverse event	24 (7.3)	26 (7.9)	30 (8.9)	19 (1.9)	99 (4.9)
Death	13 (4.0)	2 (0.6)	6 (1.8)	35 (3.5)	56 (2.8)
Failure to meet inclusion criteria	1 (0.3)	1 (0.3)	0 (0.0)	0 (0.0)	2 (0.1)
Lost to follow-up	5 (1.5)	3 (0.9)	4 (1.2)	16 (1.6)	28 (1.4)
Other	3 (0.9)	5 (1.5)	3 (0.9)	10 (1.0)	21 (1.0)
Physician decision	1 (0.3)	2 (0.6)	1 (0.3)	9 (0.9)	13 (0.6)
Protocol deviation	0 (0.0)	1 (0.3)	1 (0.3)	3 (0.3)	5 (0.2)
Withdrawal by subject	4 (1.2)	4 (1.2)	10 (3.0)	47 (4.7)	65 (3.2)
Retrieved dropouts*	**21 (6.3)**	**22 (6.7)**	**39 (11.5)**	**49 (4.9)**	**131 (6.5)**
Missed primary endpoints*	**29 (8.8)**	**20 (6.1)**	**15 (4.4)**	**88 (8.8)**	**152 (7.6)**

* Information on "Retrieved Dropouts" and "Missed Primary Endpoints" were provided by the reviewer based on the datasets: adsl and ada1c. Patients inadvertently enrolled in the study were not counted in either category.

Source: Table GPGM.8.2, CSR & reviewer's analysis.

Table 6: Patient Disposition, All Randomized Population, SURPASS-5

SURPASS-5 (GPGI)	**TZP 5mg (N = 116)**	**TZP 5mg (N = 119)**	**TZP 5mg (N = 120)**	**Placebo (N = 120)**	**Total (N = 475)**
Randomized and treated	**116 (100.0)**	**119 (100.0)**	**120 (100.0)**	**120 (100.0)**	**475 (100.0)**
Inadvertent enrollment	0 (0.0)	1 (0.8)	2 (1.7)	1 (0.8)	4 (0.8)
Completed study	109 (94.0)	115 (96.6)	110 (91.7)	117 (97.5)	451 (94.9)
Completed treatment	105 (90.5)	105 (88.2)	98 (81.7)	116 (96.7)	424 (89.3)
Without rescue medication	105 (90.5)	105 (88.2)	97 (80.8)	111 (92.5)	418 (88.0)
With rescue medication	0 (0.0)	0 (0.0)	1 (0.8)	5 (4.2)	6 (1.3)
Discontinued study prior to the primary endpoint visit	**7 (6.0)**	**4 (3.4)**	**10 (8.3)**	**2 (1.7)**	**23 (4.8)**
Adverse event	3 (2.6)	0 (0.0)	2 (1.7)	0 (0.0)	5 (1.1)
Lost to follow-up	0 (0.0)	0 (0.0)	1 (0.8)	0 (0.0)	1 (0.2)
Other	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Physician decision	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Protocol deviation	0 (0.0)	1 (0.8)	2 (1.7)	1 (0.8)	4 (0.8)
Withdrawal by subject	4 (3.4)	3 (2.5)	5 (4.2)	1 (0.8)	13 (2.7)
Discontinued treatment	**11 (9.5)**	**14 (11.8)**	**22 (18.3)**	**4 (3.3)**	**51 (10.7)**
Adverse event	7 (6.0)	10 (8.4)	13 (10.8)	3 (2.5)	33 (6.9)
Lost to follow-up	0 (0.0)	0 (0.0)	1 (0.8)	0 (0.0)	1 (0.2)
Other	0 (0.0)	0 (0.0)	1 (0.8)	0 (0.0)	1 (0.2)
Physician decision	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Protocol deviation	0 (0.0)	1 (0.8)	2 (1.7)	1 (0.8)	4 (0.8)
Withdrawal by subject	4 (3.4)	3 (2.5)	5 (4.2)	0 (0.0)	12 (2.5)
Retrieved dropouts*	**4 (3.4)**	**10 (8.4)**	**12 (10.0)**	**1 (0.8)**	**27 (5.7)**
Missed primary endpoints*	**7 (6.0)**	**3 (2.5)**	**8 (6.7)**	**2 (1.7)**	**20 (4.2)**

* Information on "Retrieved Dropouts" and "Missed Primary Endpoints" were provided by the reviewer based on the datasets: adsl and ada1c. Patients inadvertently enrolled in the study were not counted in either category.

Source: Table GPGI.8.2, CSR & reviewer's analysis.

Reference ID: 4938321

The summaries of patient demographics and baseline characteristics follow next. For each trial, the demographics and baseline characteristics were well-balanced across the study arms. Of note, the five trials focused on subjects at different stages and of various severities of the disease. In particular, SURPASS-1 studied subjects at relatively early stages (duration of diabetes 4.7 years), and with relatively low baseline HbA1c levels (7.94%), whereas SURPASS-4 and -5 targeted at subjects at later stages (duration of diabetes 11.78 years and 13.30 years) and with high glucose levels (8.52% and 8.31%).

Table 7: Patient Demographics and Baseline Characteristics, All Randomized Patients, SURPASS-1

SURPASS-1 (GPGK)	**TZP 5mg (N = 121)**	**TZP 10mg (N = 121)**	**TZP 15mg (N = 121)**	**Placebo (N = 115)**	**Total (N = 478)**
Age (years)					
Mean (SD)	54.1 (11.9)	55.8 (10.4)	52.9 (12.3)	53.6 (12.8)	54.1 (11.9)
Median (min, max)	55.0 (23, 83)	56.0 (30, 77)	53.0 (27, 75)	54.0 (18, 88)	54.5 (18, 88)
< 65 years old n (%)	94 (77.7)	91 (75.2)	99 (81.8)	89 (77.4)	373 (78.0)
≥ 65 years old n (%)	27 (22.3)	30 (24.8)	22 (18.2)	26 (22.6)	105 (22.0)
Sex					
Female n (%)	65 (53.7)	49 (40.5)	58 (47.9)	59 (51.3)	231 (48.3)
Male n (%)	56 (46.3)	72 (59.5)	63 (52.1)	56 (48.7)	247 (51.7)
Ethnicity					
Hispanic or Latino	50 (41.3)	54 (44.6)	55 (45.5)	48 (41.7)	207 (43.3)
No Hispanic or Latino	48 (39.7)	49 (40.5)	42 (34.7)	45 (39.1)	184 (38.5)
Not reported	23 (19.0)	18 (14.9)	24 (19.8)	22 (19.1)	87 (18.2)
Race					
American Indian or Alaska native	31 (25.6)	31 (25.6)	30 (24.8)	26 (22.6)	118 (24.7)
Asian	45 (37.2)	43 (35.5)	42 (34.7)	38 (33.0)	168 (35.1)
Black	7 (5.8)	4 (3.3)	6 (5.0)	5 (4.3)	22 (4.6)
White	38 (31.4)	43 (35.5)	43 (35.5)	46 (40.0)	170 (35.6)
Country n (%)					
India	18 (14.9)	20 (16.5)	18 (14.9)	17 (14.8)	73 (15.3)
Japan	23 (19.0)	22 (18.2)	23 (33.1)	21 (18.3)	89 (18.6)
Mexico	42 (34.7)	40 (33.1)	42 (34.7)	40 (34.8)	164 (34.3)
United States	38 (31.4)	39 (32.2)	38 (31.4)	37 (32.2)	152 (31.8)
Duration of diabetes (years)					
Mean (SD)	4.6 (5.1)	4.9 (5.6)	4.8 (5.0)	4.5 (5.9)	4.7 (5.4)
Median (min, max)	2.8 (0.0, 23.9)	2.9 (0.0, 30.9)	2.9 (0.0, 22.8)	2.6 (0.0, 32.8)	2.8 (0.0, 32.8)
HbA1c (%) at baseline					
Mean (SD)	7.97 (0.84)	7.90 (0.78)	7.85 (1.02)	8.05 (0.80)	7.94 (0.87)
Median (min, max)	7.90 (6.1, 10.7)	7.80 (6.4, 10.6)	7.70 (5.2, 11.5)	7.90 (6.7, 10.8)	7.80 (5.2, 11.5)
Prior use of OAMs					
Yes n (%)	66 (54.5)	68 (56.2)	65 (53.7)	60 (52.2)	259 (54.2)
No n (%)	55 (45.5)	53 (43.8)	56 (46.3)	55 (47.8)	219 (45.8)
Body mass index (kg/m^2)					
Mean (SD)	32.2 (7.0)	32.2 (7.6)	31.5 (5.5)	31.7 (6.1)	31.9 (6.6)
Median (min, max)	30.1 (21.9, 60.5)	29.8 (22.7, 68.3)	30.6 (23.0, 46.7)	30.8 (21.6, 48.8)	30.4 (21.6, 68.3)

Source: Table GPGK.8.3, CSR

Reference ID: 4938321

Table 8: Patient Demographics and Baseline Characteristics, All Randomized Patients, SURPASS-2

SURPASS-2 (GPGL)	TZP 5mg (N = 470)	TZP 10mg (N = 469)	TZP 15mg (N = 470)	SEMA (N = 469)	Total (N = 1878)
Age (years)					
Mean (SD)	56.3 (10.0)	57.2 (10.5)	55.9 (10.4)	56.9 (10.8)	56.6 (10.4)
Median (min, max)	56.0 (24, 82)	58.0 (23, 91)	56.0 (21, 81)	58.0 (28, 82)	57.0 (21, 91)
< 65 years old n (%)	360 (76.6)	348 (74.2)	366 (77.9)	346 (73.8)	1420 (75.6)
≥ 65 years old n (%)	110 (23.4)	121 (25.8)	104 (22.1)	123 (26.2)	458 (24.4)
Sex					
Female n (%)	265 (56.4)	231 (49.3)	256 (54.5)	244 (52.0)	996 (53.0)
Male n (%)	205 (43.6)	238 (50.7)	214 (45.5)	255 (48.0)	882 (47.0)
Ethnicity					
Hispanic or Latino	325 (69.1)	322 (68.7)	334 (71.1)	336 (71.6)	1317 (70.1)
Not Hispanic or Latino	145 (30.9)	147 (31.3)	136 (28.9)	133 (28.4)	561 (29.9)
Race					
American Indian or Alaska native	53 (11.3)	53 (11.3)	57 (12.1)	45 (9.6)	208 (11.1)
Asian	6 (1.3)	11 (2.3)	5 (1.1)	3 (0.6)	12 (0.6)
Black	28 (6.0)	21 (4.5)	15 (3.2)	15 (3.2)	79 (4.2)
Multiple	1 (0.2)	8 (1.7)	0	3 (0.6)	12 (0.6)
Native Hawaiian or other pacific islander	0	0	1 (0.2)	2 (0.4)	3 (0.2)
White	382 (81.6)	376 (80.2)	392 (83.4)	401 (85.5)	1551 (82.6)
Country n (%)					
Argentina	158 (33.6)	160 (34.1)	161 (34.3)	161 (34.3)	640 (34.1)
Australia	12 (2.6)	11 (2.3)	12 (2.6)	11 (2.3)	46 (2.4)
Brazil	37 (7.9)	37 (7.9)	36 (7.7)	37 (7.9)	147 (7.8)
Canada	15 (3.2)	15 (3.2)	14 (3.0)	15 (3.2)	59 (3.1)
Israel	22 (4.7)	22 (4.7)	22 (4.7)	21 (4.5)	87 (4.6)
Mexico	89 (18.9)	87 (18.6)	88 (18.7)	88 (18.8)	352 (18.7)
United Kingdom	18 (3.8)	18 (3.8)	18 (3.8)	18 (3.8)	72 (3.8)
United States	119 (25.3)	119 (25.4)	119 (25.3)	118 (25.2)	475 (25.3)
Duration of diabetes (years)					
Mean (SD)	9.1 (7.2)	8.4 (5.9)	8.7 (6.9)	8.3 (5.8)	8.6 (6.5)
Median (min, max)	7.1 (0.3, 39.8)	7.1 (0.3, 30.7)	6.8 (0.3, 42.0)	7.1 (0.3, 35.1)	7.1 (0.3, 42.0)
HbA1c (%) at baseline					
Mean (SD)	8.32 (1.08)	8.30 (1.02)	8.26 (1.00)	8.25 (1.01)	8.28 (1.03)
Median (min, max)	8.20 (5.9, 12.2)	8.10 (5.7, 12.2)	8.10 (5.6, 11.3)	8.10 (5.6, 11.4)	8.10 (5.6, 12.2)
Baseline metformin use n					
Yes	470	469	470	469	1878
No	0	0	0	0	0
Body mass index (kg/m^2)					
Mean (SD)	33.8 (6.9)	34.3 (6.6)	34.5 (7.1)	34.2 (7.2)	34.2 (6.9)
Median (min, max)	32.46 (24.58, 85.61)	33.11 (24.18, 66.51)	32.91 (22.68, 76.74)	32.77 (24.44, 89.32)	32.795 (22.68, 89.32)

Source: Table GPGL.8.5, CSR

Reference ID: 4938321

Table 9: Patient Demographics and Baseline Characteristics, All Randomized Patients, SURPASS-3

SURPASS-3 (GPGH)	TZP 5mg (N = 358)	TZP 10mg (N = 360)	TZP 15mg (N = 359)	Degludec (N = 360)	Total (N = 1437)
Age (years)					
Mean (SD)	57.2 (10.1)	57.4 (9.7)	57.5 (10.2)	57.5 (10.1)	57.4 (10.0)
Median (min, max)	58.0 (22, 81)	57.0 (30, 82)	58.0 (24, 84)	58.0 (22, 81)	58.0 (22, 84)
< 65 years old n (%)	270 (75.4)	269 (74.7)	254 (70.8)	265 (73.6)	1058 (73.6)
≥ 65 years old n (%)	88 (24.6)	91 (25.3)	105 (29.2)	95 (26.4)	379 (26.4)
Sex					
Female n (%)	158 (44.1)	165 (45.8)	165 (46.0)	147 (40.8)	635 (44.2)
Male n (%)	200 (55.9)	195 (54.2)	194 (54.0)	213 (59.2)	802 (55.8)
Ethnicity					
Hispanic or Latino	109 (30.4)	108 (30.0)	96 (26.7)	108 (30.0)	421 (29.3)
Not Hispanic or Latino	246 (68.7)	252 (70.0)	259 (72.1)	252 (70.0)	1009 (70.2)
Not reported	3 (0.8)	0	4 (1.1)	0	7 (0.5)
Race					
American Indian or Alaska native	0	1 (0.3)	1 (0.3)	2 (0.6)	4 (0.3)
Asian	20 (5.6)	19 (5.3)	20 (5.6)	17 (4.7)	76 (5.3)
Black	13 (3.6)	12 (3.3)	8 (2.2)	11 (3.1)	44 (3.1)
Multiple	1 (0.3)	0	1 (0.3)	0	2 (0.1)
Native Hawaiian or other pacific islander	1 (0.3)	0	2 (0.6)	1 (0.3)	4 (0.3)
White	323 (90.2)	328 (91.1)	327 (91.1)	329 (91.4)	1307 (91.0)
Country n (%)					
Argentina	57 (15.9)	54 (15.0)	56 (15.6)	57 (15.8)	224 (15.6)
Australia	6 (1.7)	8 (2.2)	4 (1.1)	7 (1.9)	25 (1.7)
Greece	33 (9.2)	35 (9.7)	35 (9.7)	34 (9.4)	137 (9.5)
Hungary	36 (10.1)	37 (10.3)	37 (10.3)	35 (9.7)	145 (10.1)
Italy	6 (1.7)	7 (1.9)	8 (2.2)	8 (2.2)	29 (2.0)
Poland	33 (9.2)	33 (9.2)	33 (9.2)	33 (9.2)	132 (9.2)
Romania	54 (15.1)	54 (15.0)	53 (14.8)	53 (14.7)	214 (14.9)
South Korea	9 (2.5)	10 (2.8)	8 (2.2)	7 (1.9)	34 (2.4)
Spain	21 (5.9)	21 (5.8)	22 (6.1)	22 (6.1)	86 (6.0)
Taiwan	9 (2.5)	8 (2.2)	10 (2.8)	9 (2.5)	36 (2.5)
Ukraine	11 (3.1)	11 (3.1)	12 (3.3)	11 (3.1)	45 (3.1)
United States	83 (23.2)	82 (22.8)	81 (22.6)	84 (23.3)	330 (23.0)
Duration of diabetes (years)					
Mean (SD)	8.47 (5.83)	8.43 (6.59)	8.52 (6.47)	8.12 (6.04)	8.38 (6.24)
Median (min, max)	7.69 (0.1, 41.5)	6.75 (0.2, 59.7)	7.44 (0.0, 51.9)	6.68 (0.0, 40.6)	6.24 (0.0, 59.7)
HbA1c (%) at baseline					
Mean (SD)	8.17 (0.89)	8.18 (0.89)	8.21 (0.94)	8.12 (0.94)	8.17 (0.91)
Median (min, max)	8.00 (5.8, 11.2)	8.00 (6.5, 11.5)	8.00 (6.5, 10.7)	7.90 (4.9, 10.9)	8.00 (4.9, 11.5)
Use of OAM n (%)					
Metformin	246 (68.7)	242 (67.2)	247 (68.8)	244 (67.8)	979 (68.1)
Metformin + SGLT-2i	112 (31.3)	118 (32.8)	112 (31.2)	116 (32.2)	458 (31.9)
Body mass index (kg/m^2)					
Mean (SD)	33.58 (5.87)	33.41 (6.21)	33.68 (6.11)	33.42 (6.06)	33.52 (6.06)
Median (min, max)	32.70 (23.8, 62.9)	32.38 (21.5, 67.4)	32.90 (22.9, 64.6)	32.17 (21.6, 63.4)	32.56 (21.5, 67.4)

Source: Table GPGH.8.4, CSR

Reference ID: 4938321

Table 10: Patient Demographics and Baseline Characteristics, All Randomized Patients, SURPASS-4

SURPASS-4 (GPGM)	TZP 5mg (N = 329)	TZP 10mg (N = 328)	TZP 15mg (N = 338)	Insulin Glargine (N = 1000)	Total (N = 1995)
Age (years)					
Mean (SD)	62.9 (8.6)	63.7 (8.7)	63.7 (8.6)	63.8 (8.5)	63.6 (8.6)
Median (min, max)	64.0 (37, 86)	64.0 (32, 87)	64.0 (39, 91)	64.0 (36, 87)	64.0 (32, 91)
< 65 years old n (%)	185 (56.2)	166 (50.6)	180 (53.3)	516 (51.6)	1047 (52.5)
≥ 65 years old n (%)	144 (43.8)	162 (49.4)	158 (46.7)	484 (48.4)	948 (47.5)
Sex					
Female n (%)	131(39.8)	119 (36.3)	135 (39.9)	364 (36.4)	749 (37.5)
Male n (%)	198 (60.2)	209 (63.7)	203 (60.1)	636 (63.6)	1246 (62.5)
Ethnicity					
Hispanic or Latino	154 (46.8)	164 (50.0)	155 (45.9)	477 (47.7)	950 (47.6)
Not Hispanic or Latino	172 (52.3)	162 (49.4)	179 (53.0)	517 (51.7)	1030 (51.6)
Not reported	3 (0.9)	2 (0.6)	4 (1.2)	6 (0.6)	15 (0.8)
Race					
American Indian or Alaska native	32 (9.8)	30 (9.1)	26 (7.7)	85 (8.5)	173 (8.7)
Asian	15 (4.6)	16 (4.9)	8 (2.4)	31 (3.1)	70 (3.5)
Black	13 (4.0)	17 (5.2)	11 (3.3)	32 (3.2)	73 (3.7)
Multiple	8 (2.4)	6 (1.8)	5 (1.5)	24 (2.4)	43 (2.2)
Native Hawaiian or other pacific islander	0	0	2 (0.6)	1 (0.1)	3 (0.2)
White	260 (79.3)	259 (79.0)	285 (84.6)	825 (82.7)	1629 (81.8)
Missing	1	0	1	2	4
Country n (%)					
Argentina	80 (24.3)	80 (24.4)	82 (24.3)	242 (24.2)	484 (24.3)
Australia	3 (0.9)	3 (0.9)	4 (1.2)	9 (0.9)	19 (1.0)
Brazil	42 (12.8)	41 (12.5)	42 (12.4)	125 (12.5)	250 (12.5)
Canada	8 (2.4)	9 (2.7)	7 (2.1)	27 (2.7)	51 (2.6)
Greece	7 (2.1)	5 (1.5)	7 (2.1)	19 (1.9)	38 (1.9)
Israel	4 (1.2)	5 (1.5)	5 (1.5)	15 (1.5)	29 (1.5)
Mexico	27 (8.2)	27 (8.2)	26 (7.7)	79 (7.9)	159 (8.0)
Poland	27 (8.2)	25 (7.6)	26 (7.7)	77 (7.7)	155 (7.8)
Romania	13 (4.0)	14 (4.3)	14 (4.1)	41 (4.1)	82 (4.1)
Russian	14 (4.3)	14 (4.3)	15 (4.4)	45 (4.5)	88 (4.4)
Federation Slovakia	33 (10.0)	32 (9.8)	34 (10.1)	100 (10.0)	199 (10.0)
Spain	11 (3.3)	14 (4.3)	14 (4.1)	38 (3.8)	77 (3.9)
Taiwan	5 (1.5)	5 (1.5)	5 (1.5)	15 (1.5)	30 (1.5)
United States	55 (16.7)	54 (16.5)	57 (16.9)	168 (16.8)	334 (16.7)
Duration of diabetes (years)					
Mean (SD)	11.14 (7.08)	11.96 (7.45)	11.48 (7.54)	12.03 (7.66)	11.78 (7.51)
Median (min, max)	9.75 (0.3, 45.0)	10.64 (0.3, 48.7)	10.42 (0.3, 39.3)	10.67 (0.3, 47.6)	10.53 (0.3, 48.7)
HbA1c (%) at baseline					
Mean (SD)	8.52 (0.84)	8.59 (0.91)	8.52 (0.98)	8.50 (0.85)	8.52 (0.88)
Median (min, max)	8.30 (6.1, 11.3)	8.50 (6.7, 11.2)	8.40 (6.0, 15.8)	8.40 (5.5, 12.2)	8.40 (5.5, 15.8)
Use of SGLT-2i n(%)					
Yes	78 (23.7)	81 (24.7)	86 (25.4)	256 (25.6)	501 (25.1)
No	251 (76.3)	247 (75.3)	252 (74.6)	744 (74.4)	1494 (74.9)
Body mass index (kg/m^2)					
Mean (SD)	32.64 (6.06)	32.81 (5.51)	32.50 (5.02)	32.45 (5.55)	32.55 (5.54)
Median (min, max)	31.65 (23.59, 67.93)	32.22 (22.10, 52.69)	31.72 (23.50, 55.90)	31.53 (21.73, 61.47)	31.68 (21.73, 67.93)

Source: Table GPGM.8.3, CSR

Reference ID: 4938321

Table 11: Patient Demographics and Baseline Characteristics, All Randomized Patients, SURPASS-5

SURPASS-5 (GPGI)	TZP 5mg (N = 116)	TZP 10mg (N = 119)	TZP 15mg (N = 120)	Placebo (N = 120)	Total (N = 475)
Age (years)					
Mean (SD)	61.5 (9.8)	60.4 (10.2)	60.5 (9.9)	60.0 (9.6)	60.6 (9.9)
Median (min, max)	64 (30, 82)	61.0 (30, 80)	61.5 (27, 79)	60.0 (35, 83)	61.0 (27, 83)
< 65 years old n (%)	62 (53.4)	70 (58.8)	71 (59.2)	80 (66.7)	283 (59.6)
≥ 65 years old n (%)	54 (46.6)	49 (41.2)	49 (40.8)	40 (33.3)	192 (40.4)
Sex n (%)					
Female	55 (47.4)	47 (39.5)	55 (45.8)	54 (45.0)	211 (44.4)
Male	61 (52.6)	72 (60.5)	65 (54.2)	66 (55.0)	264 (55.6)
Ethnicity n(%)					
Hispanic or Latino	4 (3.4)	8 (6.7)	5 (4.2)	5 (4.2)	22 (4.6)
Not Hispanic or Latino	94 (81.0)	95 (79.8)	93 (77.5)	98 (81.7)	380 (80.0)
Not reported	18 (15.5)	16 (13.4)	22 (18.3)	17 (14.2)	73 (15.4)
Race n (%)					
American Indian or Alaska native	0	1 (0.8)	1 (0.8)	0	2 (0.4)
Asian	20 (17.2)	21 (17.6)	22 (18.3)	22 (18.3)	85 (17.9)
Black	1 (0.9)	2 (1.7)	3 (2.5)	0	6 (1.3)
Multiple	0	1 (0.8)	0	1 (0.8)	2 (0.4)
White	95 (81.9)	94 (79.0)	94 (78.3)	97 (80.8)	380 (80.0)
Country n (%)					
Czech Republic	24 (20.7)	24 (20.2)	23 (19.2)	23 (19.2)	94 (19.8)
Germany	32 (27.6)	32 (26.9)	33 (27.5)	32 (26.7)	129 (27.2)
Japan	19 (16.4)	21 (17.6)	20 (16.7)	22 (18.3)	82 (17.3)
Poland	8 (6.9)	9 (7.6)	10 (8.3)	9 (7.5)	36 (7.6)
Slovakia	8 (6.9)	7 (5.9)	8 (6.7)	8 (6.7)	31 (6.5)
Spain	13 (11.2)	15 (12.6)	15 (12.5)	14 (11.7)	57 (12.0)
United States	12 (0.3)	11 (9.2)	11 (9.2)	12 (10.0)	46 (9.7)
Duration of diabetes (years)					
Mean (SD)	14.11 (8.08)	12.59 (6.16)	13.65 (7.50)	12.87 (7.39)	13.30 (7.31)
Median (min, max)	12.84 (0.8, 38.7)	11.42 (2.5, 30.1)	12.84 (1.1, 35.1)	11.89 (0.6, 39.7)	11.93 (0.6, 39.7)
HbA1c (%) at baseline					
Mean (SD)	8.3 (0.88)	8.36 (0.83)	8.23 (0.86)	8.37 (0.84)	8.31 (0.85)
Median (min, max)	8.10 (6.4, 11.0)	8.25 (6.5, 10.5)	8.20 (6.3, 10.5)	8.30 (6.9, 10.7)	8.20 (6.3, 11.0)
Baseline metformin use n (%)					
Yes	17 (14.7)	20 (16.8)	23 (19.2)	21 (17.5)	81 (17.1)
No	99 (85.3)	99 (83.2)	97 (80.8)	99 (82.5)	394 (82.9)
Body mass index (kg/m^2)					
Mean (SD)	33.59 (5.94)	33.35 (6.22)	33.37 (5.88)	33.23 (6.26)	33.38 (6.06)
Median (min, max)	33.05 (23.4, 53.5)	32.60 (22.7, 55.2)	33.05 (22.8, 52.4)	32.60 (22.8, 51.1)	32.80 (22.7, 55.2)

Source: Table GPGI.8.3, CSR

3.2.4 Results and Conclusions

Primary endpoint: changes in HbA1c from baseline

The analysis results for the primary endpoint based on the methods described in Section 3.2.2, were presented as follows. The results based on the applicant's analyses concurred with the results based on the reviewer's analyses.

Reference ID: 4938321

Superiority of tirzepatide on the primary endpoint has been established for all three doses in all five studies either as a primary objective in placebo-controlled studies or as a key secondary objective in non-inferiority studies (Table 12). The most prominent comparator-adjusted treatment effect was observed in the two placebo-controlled trials: SURPASS-1 and -5; whereas the least was observed in the semaglutide-controlled trial: SURPASS-2 (due to the fact that semaglutide demonstrated better treatment effect than the other active comparators). No notable dose-response relationship relative to the primary endpoint has been found in these studies.

Table 12: Analysis Results for HbA1c (%) Change from Baseline at Week 40

	SURPASS-1	**SURPASS-2**	**SURPASS-3**	**SURPASS-4**	**SURPASS-5**
Lsmean at baseline (SE)					
Comparator	8.07 (0.08)	8.25 (0.05)	8.11 (0.05)	8.50 (0.03)	8.38 (0.08)
TZP 5mg	7.97 (0.08)	8.32 (0.05)	8.17 (0.05)	8.52 (0.05)	8.30 (0.08)
TZP 10mg	7.90 (0.08)	8.30 (0.05)	8.18 (0.05)	8.60 (0.05)	8.36 (0.08)
TZP 15mg	7.85 (0.08)	8.26 (0.05)	8.21 (0.05)	8.52 (0.05)	8.22 (0.08)
Lsmean change from baseline (SE)					
Comparator	-0.09 (0.11)	-1.84 (0.05)	-1.23 (0.06)	-1.38 (0.04)	-0.88 (0.08)
TZP 5mg	-1.75 (0.10)	-2.01 (0.05)	-1.83 (0.06)	-2.09 (0.06)	-2.21 (0.08)
TZP 10mg	-1.71 (0.11)	-2.24 (0.05)	-2.02 (0.05)	-2.31 (0.06)	-2.44 (0.08)
TZP 15mg	-1.69 (0.11)	-2.30 (0.05)	-2.13 (0.05)	-2.42 (0.05)	-2.44 (0.08)
Comparator-adjusted treatment effect with 95% CI (by the applicant's analyses)					
TZP 5mg	-1.66*** (-1.96, -1.36)	-0.16 (*) (-0.29, -0.04)	-0.60*** (-0.77, -0.44)	-0.71*** (-0.85, -0.57)	-1.33*** (-1.56, -1.10)
TZP 10mg	-1.62*** (-1.92, -1.32)	-0.40*** (-0.52, -0.27)	-0.79*** (-0.96, -0.62)	-0.92*** (-1.06, -0.78)	-1.56*** (-1.78, -1.33)
TZP 15mg	-1.60*** (-1.91, -1.29)	-0.46*** (-0.59, -0.33)	-0.91*** (-1.07, -0.74)	-1.03*** (-1.16, -0.90)	-1.56*** (-1.77, -1.33)
Comparator-adjusted treatment effect with 95% CI (by the reviewer's analyses)					
TZP 5mg	-1.64*** (-1.94, -1.34)	-0.16* (-0.29, -0.04)	-0.54*** (-0.78, -0.30)	-0.76*** (-0.89, -0.64)	-1.25*** (-1.49, -1.01)
TZP 10mg	-1.60*** (-1.90, -1.30)	-0.41*** (-0.53, -0.28)	-0.73*** (-0.98, -0.48)	-0.92*** (-1.04, -0.79)	-1.53*** (-1.77, -1.30)
TZP 15mg	-1.59*** (-1.90, -1.28)	-0.48*** (-0.60, -0.36)	-0.85*** (-1.10, -0.60)	-1.02*** (-1.14, -0.89)	-1.48*** (-1.72, -1.24)

P-values for superiority: *p-value < 0.05, *** p-value < 0.001 vs placebo or active comparator.
(*) achieved statistical significance after multiplicity adjustment

Source: The analysis results produced by the applicant and by the reviewer were based on the datasets: ada1c and adsl.

In addition, sensitivity analyses based on the return-to-baseline imputation (as specified in Section 3.2.2) yielded similar estimates of the treatment effects as the primary analysis, and hence successfully demonstrated the robustness of the primary analysis results. Details of the sensitivity analysis results can be found in Table 13, as below.

Reference ID: 4938321

Table 13: Sensitivity Analyses of the Primary Endpoint Based on the return-to-baseline Method

	SURPASS-1	SURPASS-2	SURPASS-3	SURPASS-4	SURPASS-5
Lsmean change from baseline (SE)					
Comparator	-0.09 (0.11)	-1.84 (0.05)	-1.23 (0.06)	-1.38 (0.04)	-0.88 (0.08)
TZP 5mg	-1.75 (0.10)	-2.01 (0.05)	-1.83 (0.06)	-2.09 (0.06)	-2.21 (0.08)
TZP 10mg	-1.71 (0.11)	-2.24 (0.05)	-2.02 (0.05)	-2.31 (0.06)	-2.44 (0.08)
TZP 15mg	-1.69 (0.11)	-2.30 (0.05)	-2.13 (0.05)	-2.42 (0.05)	-2.44 (0.08)
Comparator-adjusted treatment effect with 95% CI					
TZP 5mg	-1.67*** (-1.96, -1.38)	-0.18* (-0.31, -0.04)	-0.54*** (-0.69, -0.39)	-0.69*** (-0.83, -0.56)	-1.19*** (-1.45, -0.94)
TZP 10mg	-1.62*** (-1.91, -1.33)	-0.40*** (-0.53, -0.26)	-0.69*** (-0.84, -0.54)	-0.89*** (-1.03, -0.76)	-1.52*** (-1.77, -1.27)
TZP 15mg	-1.61*** (-1.91, -1.32)	-0.47*** (-0.61, -0.34)	-0.87*** (-1.02, -0.72)	-1.06*** (-1.19, -0.92)	-1.44*** (-1.69, -1.19)

P-values for superiority: *p-value < 0.05, *** p-value < 0.001 vs placebo or active comparator.

Source: The analysis results were produced by the reviewer based on datasets: ada1c and adsl.

Key secondary endpoints

Analysis results for the key secondary endpoints, based on the methods specified in Section 3.2.2 were presented next. For each study, superiority was established for TZP 5mg, 10mg and 15mg on both the primary and the key secondary endpoints, with the only exception of TZP 5mg compared to semaglutide (SURPASS-2) on incidence of HbA1c < 7%.

Table 14: Analysis Results for Weight (kg) Change from Baseline

	SURPASS-1	SURPASS-2	SURPASS-3	SURPASS-4	SURPASS-5
Lsmean at baseline (SE)					
Comparator	84.5 (1.9)	93.7 (1.0)	94.0 (1.1)	90.2 (0.6)	94.2 (2.0)
TZP 5mg	87.0 (1.8)	92.5 (1.0)	94.4 (1.1)	90.3 (1.0)	95.8 (2.0)
TZP 10mg	86.2 (1.8)	94.8 (1.0)	93.8 (1.1)	90.6 (1.0)	94.6 (2.0)
TZP 15mg	85.5 (1.8)	93.8 (1.0)	94.9 (1.1)	90.0 (0.6)	96.0 (2.0)
Lsmean change from baseline (SE)					
Comparator	-1.0 (0.5)	-5.7 (0.3)	1.9 (0.4)	1.7 (0.2)	1.6 (0.6)
TZP 5mg	-6.3 (0.5)	-7.6 (0.3)	-7.0 (0.4)	-6.4 (0.4)	-5.4 (0.6)
TZP 10mg	-7.0 (0.5)	-9.3 (0.3)	-9.6 (0.4)	-8.9 (0.4)	-7.5 (0.6)
TZP 15mg	-7.8 (0.5)	-11.2 (0.3)	-11.3 (0.4)	-10.6 (0.3)	-8.8 (0.6)
Comparator-adjusted treatment effect with 95% CI					
TZP 5mg	-5.3 *** (-6.8, -3.9)	-1.9*** (-2.8, -1.0)	-8.9*** (-10.0, -7.8)	-8.1*** (-8.9, -7.3)	-7.1*** (-8.7, -5.4)
TZP 10mg	-6.0 *** (-7.4, -4.6)	-3.6*** (-4.5, -2.7)	-11.5*** (-12.6, -10.4)	-10.6*** (-11.4, -9.8)	-9.1*** (-10.7, -7.5)
TZP 15mg	-6.9 *** (-8.3, -5.4)	-5.5*** (-6.4, -4.6)	-13.2*** (-14.3, -12.1)	-12.2*** (-13.0, -11.5)	-10.5*** (-12.1, -8.8)

P-values for superiority: *p-value < 0.05, *** p-value < 0.001 vs placebo or active comparator.

Source: The analysis results were produced by the applicant based on datasets: advs and adsl, and have been verified by the reviewer.

Reference ID: 4938321

Table 15: Analysis Results for FSG (mg/dL) Change from Baseline[8]

	SURPASS-1	SURPASS-2	SURPASS-3	SURPASS-4	SURPASS-5
Mean at baseline (SE)					
Comparator	155.2 (3.8)	171.2 (2.4)	166.6 (2.4)	168.4 (1.6)	164.4 (4.8)
TZP 5mg	153.7 (3.6)	173.8 (2.4)	171.7 (2.4)	172.3 (2.8)	162.9 (4.8)
TZP 10mg	152.6 (3.6)	174.2 (2.4)	170.4 (2.4)	175.7 (2.8)	162.6 (4.8)
TZP 15mg	155.2 (3.8)	172.4 (2.4)	168.4 (2.4)	174.1 (2.8)	160.4 (4.8)
Lsmean change from baseline (SE)					
Comparator	3.7 (4.4)	-49.4 (1.6)	-50.5 (2.5)	-48.8 (1.7)	-39.2 (2.7)
TZP 5mg	-39.6 (4.0)	-54.8 (1.5)	-47.0 (2.0)	-44.3 (2.6)	-58.2 (2.8)
TZP 10mg	-39.8 (4.1)	-59.1 (1.6)	-50.1 (21)	-50.3 (2.4)	-64.0 (2.7)
TZP 15mg	-38.6 (4.4)	-60.2 (1.5)	-54.2 (1.9)	-54.5 (2.3)	-62.6 (2.8)
Comparator-adjusted treatment effect with 95% CI					
TZP 5mg	-43.2*** (-54.8, -31.6)	-5.4[ǂ] (-9.7, -1.1)	3.5 (-2.8, 9.7)	4.5 (-1.7, 10.8)	-19.0*** (-26.6, -11.4)
TZP 10mg	-43.4*** (-55.1, -31.7)	-9.7[ǂǂ] (-14.1, -5.2)	0.3 (-6.0, 6.6)	-1.5 (-7.3, 4.3)	-24.9*** (-32.3, -17.4)
TZP 15mg	-42.3*** (-54.4, -30.3)	-10.8[ǂǂ] (-15.1, -6.5)	-3.8 (-9.9, 2.4)	-5.7[ǂ] (-11.3, -0.1)	-23.4*** (-31.0, -15.8)

P-values for superiority: *p-value < 0.05, *** p-value < 0.001 vs placebo or active comparator.
P-values not controlled for type 1 error: [ǂ] P-value < 0.05, [ǂǂ] p-value < 0.001 vs placebo or active comparator.

Source: The analysis results were produced by the applicant based on datasets: adlbcn and adsl, and have been verified by the reviewer.

Table 16: Analysis Results for Incidence of HbA1c < 7%[9]

	SURPASS-1	SURPASS-2	SURPASS-3	SURPASS-4	SURPASS-5
Percentage of subjects achieving HbA1c < 7% (SE)					
Comparator	23.0 (4.2)	79.0 (2.0)	58.0 (2.7)	48.8 (1.8)	34.5 (4.4)
TZP 5mg	81.8 (3.7)	82.0 (1.8)	79.2 (2.3)	75.1 (2.7)	87.3 (3.3)
TZP 10mg	84.5 (3.5)	85.6 (1.7)	81.5 (2.4)	82.9 (2.3)	89.6 (2.9)
TZP 15mg	78.3 (4.1)	86.2 (1.7)	83.5 (2.0)	84.9 (2.0)	84.7 (3.6)
Odds ratio (treatment/comparator) (95% CI)					
TZP 5mg	17.4*** (8.5, 35.7)	1.3 (0.9, 1.8)	3.1*** (2.1, 4.4)	3.5*** (2.5, 4.8)	14.7*** (7.0, 30.6)
TZP 10mg	21.2*** (10.1, 44.4)	1.7** (1.2, 2.5)	3.6*** (2.4, 5.4)	6.0*** (4.2, 8.7)	19.5*** (9.2, 41.3)
TZP 15mg	13.4*** (6.7, 26.8)	1.7** (1.2, 2.5)	4.2*** (2.9, 6.2)	6.8*** (4.8, 9.7)	11.5*** (5.6, 23.3)

P-values for superiority: **p-value < 0.01, *** p-value < 0.001 vs placebo or active comparator.

Source: The analysis results were produced by the applicant based on datasets: ada1c and adsl, and have been verified by the reviewer.

[8] FSG was a key secondary endpoint only for the two placebo-controlled trials: SURPASS-1 and SURPASS-5.
[9] TZP 5mg failed to demonstrate superiority to semaglutide, because more subjects treated with semaglutide achieved HbA1c < 7% than subjects on placebo/insulin control, while the proportion of subjects achieved HbA1c <7% in TZP 5mg was similar across all studies.

Reference ID: 4938321

Efficacy Conclusion

In summary, significantly large treatment effects regarding both the primary endpoint and the key secondary endpoints have been observed in all three doses of tirzepatide applied either as a monotherapy or combined with other background therapies. The comparisons were made relative to placebo, semaglutide, insulin glargine, and insulin degludec, and among participants with different disease durations and on various background therapies. The robustness of the primary analysis results was supported by sensitivity analyses.

3.3 Evaluation of Safety

The population for safety assessment was identical to the population for the primary efficacy assessment, *i.e.*, the mITT population. The safety analysis set was defined as all available data obtained from the safety population during the planned treatment period and the safety follow-up period, regardless of initiation of new antihyperglycemic medication or adherence to study drug. Key safety endpoints included treatment-emergent adverse events, early discontinuation of study drug due to adverse events (AEs), adjudicated pancreatic AEs, incidence of allergic and hypersensitivity reactions, and occurrence of hypoglycemia episodes, *etc.* A comprehensive safety evaluation can be found in the clinical review by Dr. Frank Pucino. For this statistical review, the focus was on the incidence of clinically significant hypoglycemia (Level 2 or L2) or severe hypoglycemia (Level 3 or L3)[10].

As pre-specified in the SAPs, the incidence of hypoglycemic event was analyzed using a logistic regression with treatment and stratification factors as fixed effects. The rate of hypoglycemic episodes per patient per year was analyzed using a generalized linear mixed effects model assuming the number of hypoglycemic episodes follows a negative binomial distribution with the mean modeled using stratification factors and treatment as fixed effects. When the number of hypoglycemic events was less than 10, a description of events was provided instead. The analyses did not indicate any increased risk of L2 or L3 hypoglycemia, when tirzepatide was used as a monotherapy, or combined with other background therapies. Details of these analyses follow next.

<u>SURPASS-1</u>
No patients experienced severe hypoglycemia during the study. No patient treated with tirzepatide experienced Level 2 hypoglycemic episodes. One patient treated with placebo experienced three such events.

[10] Per American Diabetes Association definition, Level 2 hypoglycemia is featured by glucose level <54 mg/dl (3.0 mmol/L), and Level 3 hypoglycemia refers to hypoglycemia associated with severe cognitive impairment requiring external assistance for recovery.

Reference ID: 4938321

SURPASS-2, -3, -4 and -5

Table 17: Summary and Analysis of Level 2 or Level 3 Hypoglycemia Incidence and Rate (SURPASS-2, Modified Intent-to-Treat Population, Safety Analysis Set)

	Incidence			Incidence Rate		
	N	n (%)	p-value*	No. of episodes	Mean rate (SE)	Relative Rate (TZP/Control) (95% CI)ǂ
Comparator	470	2 (0.43)		2	0.0638 (0.00427)	
TZP 5mg	469	4 (0.85)	0.419	4	0.0667 (0.00543)	1.05 (0.85, 1.28)
TZP 10mg	470	1 (0.21)	0.589	2	0.0633 (0.00541)	0.99 (0.80, 1.23)
TZP 15mg	469	8 (1.70)	0.062	10	0.0722 (0.00891)	1.13 (0.86, 1.49)

Note: n = number of subjects that experienced at least 1 episode of hypoglycemia; N = number of subjects in the FAS;
* Based on a logistic regression between TZP and comparator: incidence ~ pooled country + baseline A1c group (<=8.5%, >8.5%) + treatment
ǂ Based a negative binomial model between TZP and comparator: no of episodes ~ pooled country + baseline A1c group (<=8.5%, >8.5%) + treatment, offset by exposure days.

Source Table GPGL.8.108, CSR

Table 18: Summary and Analysis of Level 2 or Level 3 Hypoglycemia Incidence and Rate (SURPASS-3, Modified Intent-to-Treat Population, Safety Analysis Set)

	Incidence			Incidence Rate		
	N	n (%)	p-value*	No. of episodes	Mean Rate (SE) ǂ	Relative Rate (TZP/Control) (95% CI)ǂ
Comparator	358	26 (7.26)		38	0.1020 (0.05678)	
TZP 5mg	356	5 (1.40)	<0.001	5	0.0137 (0.00564)	0.13 (0.04, 0.44)
TZP 10mg	360	4 (1.11)	<0.001	4	0.0108 (0.00442)	0.11 (0.03, 0.40)
TZP 15mg	359	8 (2.23)	0.002	11	0.0275 (0.01434)	0.27 (0.11, 0.66)

Note: n = number of subjects that experienced at least 1 episode of hypoglycemia; N = number of subjects in the FAS;
* Based on a logistic regression between TZP and comparator: Incidence ~ Pooled country + Baseline A1c group (<=8.5%, >8.5%) + Baseline SGLT-2i use + Treatment
ǂ Based a negative binomial model between TZP and comparator: No of episodes ~ Pooled country + Baseline A1c group (<=8.5%, >8.5%) + Baseline SGLT-2i use + Treatment, offset by exposure days.

Source Table GPGH.8.117, CSR

Reference ID: 4938321

Table 19: Summary and Analysis of Level 2 or Level 3 Hypoglycemia Incidence and Rate (SURPASS-4, Modified Intent-to-Treat Population, Safety Analysis Set)

	Incidence			Incidence Rate		
	N	n (%)	p-value*	No. of episodes	Mean Rate (SE) ǂ	Relative Rate (TZP/Control) (95% CI)ǂ
Comparator	1000	200 (20.00)		535	0.34 (0.035)	
TZP 5mg	329	31 (9.42)	<0.001	62	0.11 (0.033)	0.33 (0.18, 0.59)
TZP 10mg	328	20 (6.10)	<0.001	43	0.08 (0.020)	0.22 (0.13, 0.38)
TZP 15mg	338	29 (8.58)	<0.001	64	0.12 (0.030)	0.34 (0.20, 0.59)

Note: n = number of subjects that experienced at least 1 episode of hypoglycemia; N = number of subjects in the FAS;
* Based on a logistic regression between TZP and comparator: Incidence ~ Pooled country + Baseline A1c group (<=8.5%, >8.5%) + Baseline SGLT-2i use + Treatment
ǂ Based a negative binomial model between TZP and comparator: No of episodes ~ Pooled country + Baseline A1c group (<=8.5%, >8.5%) + Baseline SGLT-2i use + Treatment, offset by exposure days.

Source Table GPGM.8.147, CSR

Table 20: Summary and Analysis of Level 2 or Level 3 Hypoglycemia Incidence and Rate (SURPASS-5, Modified Intent-to-Treat Population, Safety Analysis Set)

	Incidence			Incidence Rate		
	N	n (%)	p-value*	No. of episodes	Mean Rate (SE) ǂ	Relative Rate (TZP/Control) (95% CI)ǂ
Comparator	120	16 (13.33)		46	0.54 (0.154)	
TZP 5mg	116	18 (15.52)	0.585	61	0.49 (0.138)	0.92(0.44, 1.89)
TZP 10mg	119	24 (20.17)	0.120	52	0.69 (0.174)	1.29 (0.62, 2.69)
TZP 15mg	120	18 (15.00)	0.616	42	0.40 (0.102)	0.75 (0.36, 1.57)

Note: n = number of subjects that experienced at least 1 episode of hypoglycemia; N = number of subjects in the FAS.
* Based on a logistic regression between TZP and comparator: Incidence ~ Pooled country + Baseline A1c group (<=8.0%, >8.0%) + Baseline metformin use + Treatment.
ǂ Based a negative binomial model between TZP and comparator: No. of episodes ~ Pooled country + Baseline A1c group (<=8.0%, >8.0%) + Baseline metformin use + Treatment, offset by exposure days.

Source Table GPGI.8.120, CSR

Safety Conclusion

The results from SURPASS-1 and SURPASS-5 (Tables 17 and 20) suggested that the incidence and the incidence rate of L2 or L3 hypoglycemia were comparable between the TZP- and the placebo-treated subjects. No notable difference was observed between the TZP-treated subjects and the semaglutide-treated subjects (Table 18). When compared to insulin-treated subjects, TZP-treated subjects exhibited significantly lower incidence and event rate of L2 and L3 hypoglycemia (Tables 18 and 19). No consistent dose-response pattern relative to L2 and L3 hypoglycemia was observed across the studies: the three doses had similar incidence and the event rates in SURPASS-1, -4 and -5, whereas the 15mg dose had higher incidence and event rate than the other two doses in SURPASS-2 and -3.

Reference ID: 4938321

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

For each of the five pivotal studies, subgroup analyses on HbA1c (%) change from baseline were conducted with respect to the patient's baseline characteristics: sex (Male or Female), region (the US or outside the US), race (Asian, Black, and White, etc.), and age (< 65 years or >= 65 years). Each analysis applied the same statistical method as the corresponding primary efficacy analysis (*i.e.*, ANCOVA models with missing primary endpoints imputed based on retrieved dropout for SURPASS-2, -3 and -4, or placebo washout for SURPASS-1 or -5). For each baseline characteristic, interactions between subgroups and treatment arms were tested.

Additionally, the Bayesian shrinkage analyses based on the sample estimates derived from the traditional subgroup analyses were performed. For each study, while estimating the treatment effect within a subgroup of a given baseline characteristic (e.g., the male subgroup), the shrinkage method borrowed information from other subgroup(s) (the female subgroup), and thus was considered a "weighted" average of the sample estimate and the overall estimate. The weights were based on the ratio of the between-subgroup variability to the within-subgroup variability. A small ratio indicated a small between-subgroup variability relative to the within-subgroup variability. Consequently, more weight was put on the overall estimate, and more shrinkage was applied.

For a given baseline characteristic with *k* subgroups, let Y_i (i = 1, ... k) be the observed sample estimate of the treatment effect in subgroup *i*. The shrinkage analysis in this review assumes the following:

- $Y_i \sim N(\mu_i, \sigma_i^2)$, where μ_i is the expected treatment effect for subgroup *i*, and σ_i^2 is the within-subgroup variance
- σ_i^2 is set to the observed variance for sample estimate
- $\mu_i \sim N(\mu, \tau^2)$, where $\mu \sim N(0, (4)^2)$, and $1/\tau^2 \sim$ Gamma (0.001, 0.001)

The last assumption stated that the expected treatment effect for all *k* subgroups share a common normal distribution centered at μ and with variance τ^2. A non-informative prior, as specified above, was applied to this normal distribution. Of note, a standard deviation of 4 was chosen for the centrality parameter μ, so that its standard deviation was approximately four times the subject-level standard deviation[11].

4.1 Gender, Race, Age, and Geographic Region

The sample estimates and the shrinkage estimates of the treatment difference with respect to HbA1c change from baseline are presented in Tables 21 through 25[12]. For each study, estimates of the treatment effects based on both methods were consistent across subgroups and with the overall treatment effect. Compared to the traditional method, the shrinkage method produced

[11] The subject-level standard deviation was estimated to be around 1 for each study based on the primary analysis results, as presented in Table 12.

[12] All the table contents were provided by the reviewer based statistical analyses on datasets: ada1c and adsl of each study.

Reference ID: 4938321

estimates closer to the overall estimates, and of less variance. In particular, for some subgroups, the narrower credible intervals derived from shrinkage analyses stay below zero, whereas the confidence intervals based on traditional analyses traversed zero (e.g., the *female* Group in SURPASS-3).

Table 21: Sample and shrinkage Estimates of Difference in HbA1c % Change from Baseline within Subgroups, mITT Population, SURPASS-1

		TZP 5mg			TZP 10mg			TZP 15mg		
Overall (95% CI)		**-1.64 (-1.94, -1.34)**			**-1.60 (-1.90, -1.30)**			**-1.59 (-1.90, -1.28)**		
		Sample (95% CI)	**Shrinkage (95% CI)**	**n**	**Sample (95% CI)**	**Shrinkage (95% CI)**	**n**	**Sample (95% CI)**	**Shrinkage (95% CI)**	**n**
Sex	Male	-1.77 (-2.18, -1.36)	-1.70 (-2.07, -1.36)	56	-1.66 (-2.06, -1.27)	-1.63 (-1.98, -1.28)	72	-1.72 (-2.13, -1.31)	-1.64 (-2.01, -1.29)	63
	Female	-1.50 (-1.92, -1.07)	-1.58 (-1.94, -1.18)	65	-1.53 (-1.99, -1.06)	-1.57 (-1.96, -1.17)	49	-1.39 (-1.85, -0.94)	-1.48 (-1.85, -1.07)	57
Region	The US	-1.62 (-2.18, -1.06)	-1.63 (-2.06, -1.19)	38	-1.62 (-2.17, -1.07)	-1.60 (-2.03, -1.17)	39	-1.33 (-1.91, -0.75)	-1.47 (-1.92, -0.93)	37
	Outside the US	-1.65 (-2.02, -1.28)	-1.64 (-1.99, -1.29)	83	-1.57 (-1.95, -1.20)	-1.58 (-1.92, -1.23)	82	-1.70 (-2.08, -1.33)	-1.63 (-2.01, -1.27)	83
Race*	American Indian/Alask -a Native	-1.53 (-2.05, -1.01)	-1.59 (-1.96, -1.21)	31	-1.46 (-2.00, -0.92)	-1.52 (-1.90, -1.12)	31	-1.73 (-2.26, -1.19)	-1.67 (-2.07, -1.30)	30
	Asian	-1.68 (-2.21, -1.16)	-1.63 (-2.01, -1.27)	45	-1.70 (-2.22, -1.17)	-1.58 (-1.99, -1.22)	43	-1.82 (-2.36, -1.29)	-1.70 (-2.12, -1.33)	41
	White	-1.65 (-2.13, -1.18)	-1.63 (-1.98, -1.27)	38	-1.48 (-1.95, -1.02)	-1.52 (-1.87, -1.15)	43	-1.44 (-1.91, -0.96)	-1.58 (-1.94, -1.17)	43
Age	≥ 65	-1.63 (-2.16, -1.11)	-1.64 (-2.04, -1.24)	27	-1.22 (-1.73, -0.70)	-1.38 (-1.82, -0.84)	30	-0.87 (-1.47, -0.28)	-1.03 (-1.66, -0.39)	21
	< 65	-1.67 (-2.02, -1.32)	-1.66 (-1.97, -1.35)	94	-1.74 (-2.09, -1.38)	-1.67 (-2.01, -1.34)	91	-1.80 (-2.15, -1.45)	-1.74 (-2.10, -1.40)	99

* Black or African American was excluded from the analyses due to insufficient sample size.

Table 22: Sample and Shrinkage Estimates of Difference in HbA1c % Change from Baseline within Subgroups, mITT Population, SURPASS-2

		TZP 5mg			TZP 10mg			TZP 15mg		
Overall (95% CI)		**-0.16 (-0.29, -0.04)**			**-0.41 (-0.53, -0.28)**			**-0.48 (-0.60, -0.36)**		
		Sample (95% CI)	**Shrinkage (95% CI)**	**n**	**Sample (95% CI)**	**Shrinkage (95% CI)**	**n**	**Sample (95% CI)**	**Shrinkage (95% CI)**	**n**
Sex	Male	-0.20 (-0.37, -0.03)	-0.19 (-0.34, -0.04)	205	-0.43 (-0.60, -0.27)	-0.42 (-0.57, -0.27)	238	-0.54 (-0.71, -0.37)	-0.52 (-0.67, -0.37)	213
	Female	-0.14 (-0.31, 0.04)	-0.15 (-0.31, 0.01)	265	-0.39 (-0.57, -0.20)	-0.40 (-0.56, -0.23)	231	-0.44 (-0.62, -0.26)	-0.46 (-0.62, -0.29)	256
Region	The US	-0.07 (-0.35, 0.21)	-0.11 (-0.33, 0.13)	119	-0.45 (-0.73, -0.16)	-0.43 (-0.66, -0.22)	119	-0.43 (-0.72, -0.15)	-0.46 (-0.67, -0.22)	119
	Outside the US	-0.19 (-0.33, -0.06)	-0.18 (-0.31, -0.05)	351	-0.40 (-0.53, -0.26)	-0.40 (-0.53, -0.28)	350	-0.50 (-0.63, -0.36)	-0.49 (-0.62, -0.37)	350
Race*	American Indian/Alask -a Native	-0.08 (-0.55, 0.38)	-0.15 (-0.43, 0.16)	53	-0.18 (-0.64, 0.29)	-0.33 (-0.61, 0.06)	53	-0.79 (-1.25, -0.34)	-0.61 (-1.04, -0.32)	57
	Asian	-0.35 (-1.82, 1.13)	-0.19 (-0.70, 0.25)	6	-0.12 (-1.43, 1.20)	-0.36 (-0.79, 0.24)	11	-0.85 (-2.46, 0.77)	-0.56 (-1.27, -0.05)	5
	Black/ African American	-0.41 (-1.14, 0.32)	-0.21 (-0.66, 0.11)	28	-0.29 (-1.04, 0.46)	-0.37 (-0.73, 0.09)	21	-0.66 (-1.46, 0.15)	-0.56 (-1.08, -0.16)	15
	White	-0.17 (-0.30, -0.04)	-0.17 (-0.29, -0.05)	382	-0.47 (-0.60, -0.34)	-0.45 (-0.58, -0.33)	376	-0.43 (-0.56, -0.30)	-0.44 (-0.57, -0.32)	391
Age	≥ 65	-0.14 (-0.40, 0.12)	-0.15 (-0.36, 0.06)	110	-0.41 (-0.67, -0.16)	-0.41 (-0.62, -0.20)	121	-0.35 (-0.62, -0.09)	-0.40 (-0.61, -0.16)	103
	< 65	-0.17 (-0.31, -0.03)	-0.17 (-0.30, -0.03)	360	-0.41 (-0.55, -0.27)	-0.41 (-0.54, -0.28)	348	-0.51 (-0.65, -0.37)	-0.49 (-0.63, -0.36)	366

* “Native Hawaiian or Other Pacific Island” and “Multiple” were excluded from the analyses due to insufficient sample sizes.

Reference ID: 4938321

Table 23: Sample and shrinkage Estimates of Difference in HbA1c % Change from Baseline within Subgroups, mITT Population, SURPASS-3

		TZP 5mg			TZP 10mg			TZP 15mg		
Overall (95% CI)		**-0.54 (-0.78, -0.30)**			**-0.73 (-0.98, -0.48)**			**-0.85 (-1.10, -0.60)**		
		Sample (95% CI)	**Shrinkage (95% CI)**	**n**	**Sample (95% CI)**	**Shrinkage (95% CI)**	**n**	**Sample (95% CI)**	**Shrinkage (95% CI)**	**n**
Sex	Male	-0.56 (-0.86, -0.25)	-0.55 (-0.82, -0.29)	200	-0.71 (-1.02, -0.40)	-0.71 (-1.03, -0.40)	195	-0.86 (-1.18, -0.55)	-0.85 (-1.17, -0.53)	193
	Female	-0.51 (-1.14, 0.11)	-0.53 (-0.98, -0.07)	158	-0.74 (-1.39, -0.09)	-0.72 (-1.19, -0.27)	165	-0.82 (-1.46, -0.18)	-0.84 (-1.31, -0.35)	165
Region	The US	-0.56 (-0.86, -0.25)	-0.56 (-0.78, -0.33)	200	-0.71 (-1.02, -0.40)	-0.74 (-0.96, -0.50)	195	-0.86 (-1.18, -0.55)	-0.88 (-1.14, -0.60)	193
	Outside the US	-0.56 (-0.74, -0.37)	-0.56 (-0.74, -0.38)	275	-0.80 (-0.99, -0.61)	-0.78 (-0.97, -0.61)	278	-0.91 (-1.10, -0.72)	-0.90 (-1.09, -0.72)	277
Race*	Asian	-0.99 (-1.72, -0.26)	-0.69 (-1.33, -0.28)	200	-0.86 (-1.63, -0.09)	-0.77 (-1.26, -0.33)	190	-1.14 (-1.82, -0.45)	-0.94 (-1.49, -0.51)	200
	Black/ African American	-0.47 (-1.52, 0.58)	-0.56 (-1.15, 0.03)	13	-0.58 (-1.67, 0.51)	-0.72 (-1.25, -0.12)	12	-0.16 (-1.45, 1.13)	-0.74 (-1.30, 0.22)	8
	White	-0.52 (-0.76, -0.28)	-0.54 (-0.73, -0.35)	323	-0.74 (-0.99, -0.48)	-0.74 (-1.00, -0.48)	328	-0.85 (-1.10, -0.60)	-0.85 (-1.11, -0.59)	326
Age	≥ 65	-0.31 (-0.66, 0.05)	-0.39 (-0.68, -0.06)	88	-0.42 (-0.78, -0.06)	-0.51 (-0.83, -0.16)	91	-0.39 (-0.74, -0.05)	-0.46 (-0.82, -0.10)	105
	< 65	-0.62 (-0.89, -0.36)	-0.56 (-0.84, -0.31)	270	-0.84 (-1.10, -0.57)	-0.78 (-1.07, -0.51)	269	-1.02 (-1.29, -0.76)	-0.97 (-1.26, -0.69)	253

* "American Indian or Alaska Native" and "Native Hawaiian or Other Pacific Islander" were excluded from the analyses due to insufficient sample sizes.

Table 24: Sample and Shrinkage Estimates of Difference in HbA1c % Change from Baseline within Subgroups, mITT Population, SURPASS-4

		TZP 5mg			TZP 10mg			TZP 15mg		
Overall (95% CI)		**-0.76 (-0.89, -0.64)**			**-0.92 (-1.04, -0.79)**			**-1.02 (-1.14, -0.89)**		
		Sample (95% CI)	**Shrinkage (95% CI)**	**n**	**Sample (95% CI)**	**Shrinkage (95% CI)**	**n**	**Sample (95% CI)**	**Shrinkage (95% CI)**	**n**
Sex	Male	-0.69 (-0.85, -0.53)	-0.71 (-0.88, -0.55)	197	-0.78 (-0.93, -0.63)	-0.80 (-0.96, -0.65)	208	-0.97 (-1.12, -0.81)	-0.99 (-1.13, -0.84)	202
	Female	-0.91 (-1.11, -0.71)	-0.87 (-1.07, -0.69)	131	-1.15 (-1.36, -0.94)	-1.11 (-1.32, -0.91)	118	-1.12 (-1.31, -0.92)	-1.08 (-1.28, -0.91)	135
Region	The US	-0.56 (-0.87, -0.25)	-0.65 (-0.91, -0.32)	550	-0.98 (-1.29, -0.67)	-0.94 (-1.22, -0.70)	530	-0.87 (-1.17, -0.57)	-0.93 (-1.15, -0.67)	570
	Outside the US	-0.81 (-0.94, -0.67)	-0.79 (0.93, -0.66)	273	-0.91 (-1.04, -0.77)	-0.91 (-1.05, -0.78)	273	-1.04 (-1.18, -0.91)	-1.02 (-1.16, -0.89)	280
Race	American Indian/Alask -a Native	-0.93 (-1.36, -0.50)	-0.81 (-1.12, -0.56)	32	-0.75 (-1.19, 0.31)	-0.92 (-1.26, -0.56)	30	-0.89 (-1.35, -0.44)	-0.97 (-1.22, -0.67)	26
	Asian	-0.54 (-1.10, 0.02)	-0.73 (-1.01, -0.39)	15	-1.16 (-1.69, -0.63)	-0.99 (-1.35, -0.73)	16	-0.82 (-1.46, -0.17)	-0.96 (-1.25, -0.60)	8
	Black/ African American	-0.85 (0.14) (-1.58, -0.13)	-0.78 (-1.15, -0.47)	13	-0.96 (-1.62, -0.30)	-0.95 (-1.30, -0.64)	17	-0.92 (-1.67, -0.18)	-0.98 (-1.29, -0.63)	11
	Multiple	-1.16 (-2.04, -0.29)	-0.81 (-1.26, -0.51)	8	-1.25 (-2.18, -0.33)	-0.98 (-1.42, -0.66)	6	-1.05 (-2.14, 0.04)	-1.00 (-1.35, -0.61)	5
	White	-0.74 (-0.88, -0.61)	-0.75 (-0.88, -0.62)	259	-0.91 (-1.05, -0.77)	-0.92 (-1.05, -0.79)	257	-1.04 (-1.17, -0.91)	-1.03 (-1.15, -0.91)	284
Age	≥ 65	-0.65 (-0.83 -0.47)	-0.69 (-0.86, -0.49)	143	-0.79 (-0.96, -0.62)	-0.83 (-1.02, -0.63)	161	-0.81 (-0.98, -0.64)	-0.85 (-1.05, -0.64)	157
	< 65	-0.86 (-1.03, -0.69)	-0.82 (-1.02, -0.64)	185	-1.04 (-1.21, -0.86)	-1.00 (-1.20, -0.81)	165	-1.18 (-1.35, -1.01)	-1.14 (-1.35, -0.95)	180

Reference ID: 4938321

Table 25: Sample and shrinkage Estimates of HbA1c % Change from Baseline within Subgroups, mITT Population, SURPASS-5

		TZP 5mg			TZP 10mg			TZP 15mg		
Overall (95% CI)		**-1.25 (-1.49, -1.01)**			**-1.53 (-1.77, -1.30)**			**-1.48 (-1.72, -1.24)**		
		Sample (95% CI)	**Shrinkage (95% CI)**	**n**	**Sample (95% CI)**	**Shrinkage (95% CI)**	**n**	**Sample (95% CI)**	**Shrinkage (95% CI)**	**n**
Sex	Male	-1.21 (-1.55, -0.87)	-1.23 (-1.51, -0.93)	61	-1.58 (-1.90, -1.26)	-1.54 (-1.84, -1.25)	72	-1.35 (-1.69, -1.02)	-1.41 (-1.71, -1.09)	63
	Female	-1.28 (-1.63, -0.93)	-1.26 (-1.56, -0.97)	55	-1.43 (-1.80, -1.07)	-1.47 (-1.76, -1.17)	46	-1.61 (-1.96, -1.26)	-1.54 (-1.86, -1.24)	55
Region	The US	-1.05 (-2.00, -0.11)	-1.17 (-1.75, -0.46)	12	-1.44 (-2.38, -0.50)	-1.49 (-2.10, -0.83)	11	-1.21 (-2.13, -0.29)	-1.38 (-1.94, -0.64)	12
	Outside the US	-1.26 (-1.51, -1.01)	-1.24 (-1.52, -0.98)	104	-1.53 (-1.78, -1.28)	-1.52 (-1.79, -1.25)	107	-1.50 (-1.75, -1.25)	-1.48 (-1.75, -1.21)	107
Race*	Asian	-1.35 (-1.82, -0.88)	-1.29 (-1.69, -0.94)	20	-1.27 (-1.73 -0.81)	-1.40 (-1.78, -0.93)	21	-1.65 (-2.10, -1.19)	-1.60 (-1.96, -1.26)	22
	White	-1.22 (-1.49, -0.96)	-1.24 (-1.49, -0.98)	95	-1.63 (-1.89, -1.36)	1.58 (-1.86, -1.32)	93	-1.52 (-1.79, -1.25)	-1.54 (-1.80, -1.28)	92
Age	≥ 65	-1.02 (-1.39, -0.64)	-1.11 (-1.46, -0.73)	54	-1.29 (-1.66, -0.92)	-1.39 (-1.73, -1.00)	49	-1.13 (-1.52, -0.75)	-1.24 (-1.61, -0.82)	48
	< 65	-1.39 (-1.71, -1.07)	-1.31 (-1.65, -1.01)	62	-1.68 (-1.99, -1.37)	-1.60 (-1.94, -1.29)	69	-1.67 (-1.98, -1.35)	-1.59 (-1.94, -1.26)	70

***** "Black or African American", "American Indian or Alaska Native" and "Multiple" were excluded from the analyses due to insufficient sample sizes.

For further illustration purpose, forest plots based on the subgroup analysis results from SURPASS-2 were demonstrated in Figures 7 through 10.

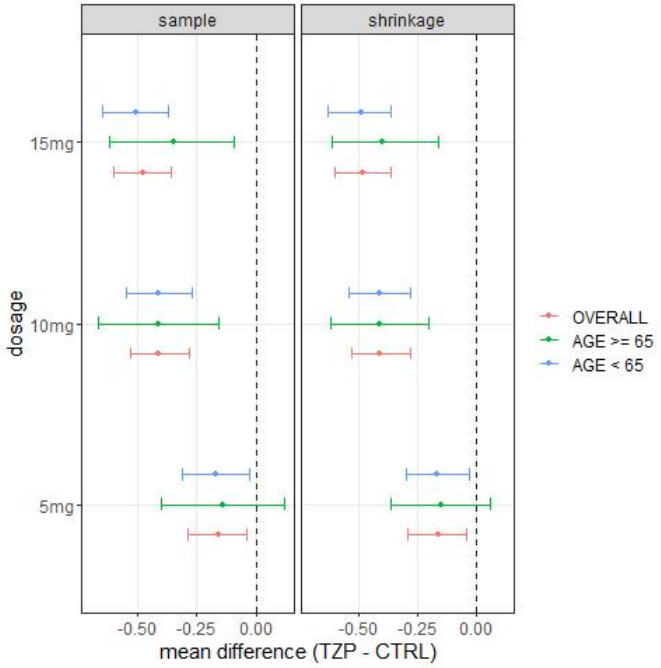


Figure 7: Subgroup Analysis for Age (< 65 Years, >= 65 Years), SURPASS-2

Reference ID: 4938321

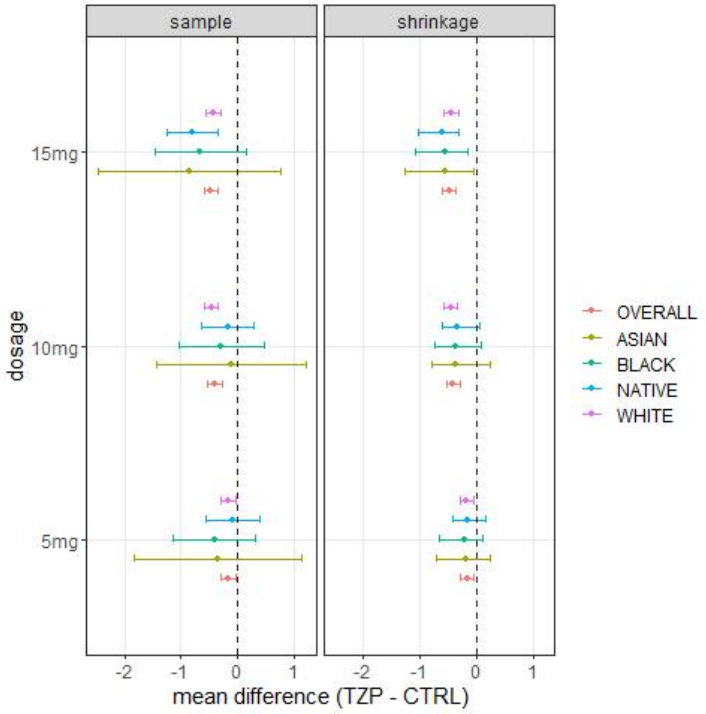


Figure 8: Subgroup Analysis for Race, SURPASS-2

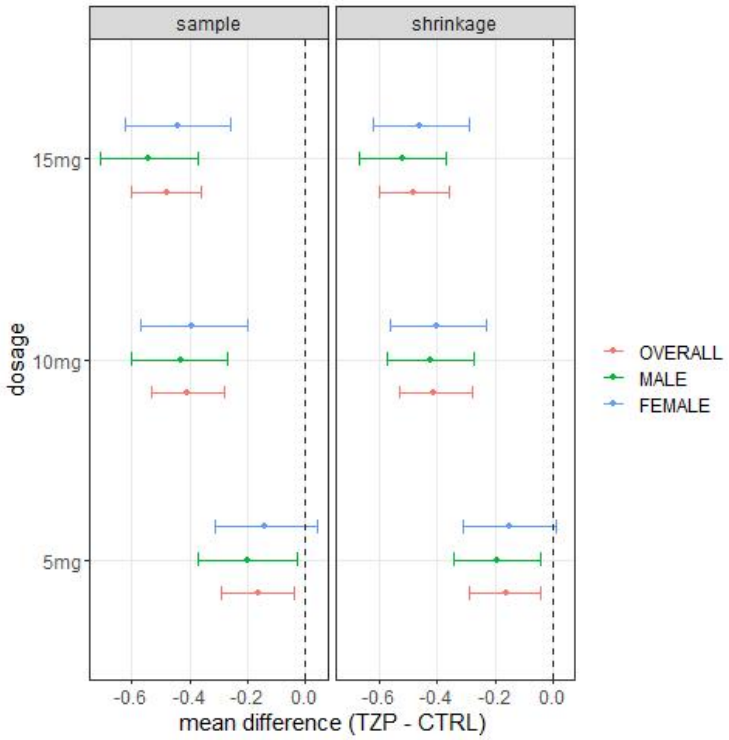


Figure 9: Subgroup Analysis for Sex, SURPASS-2

Reference ID: 4938321

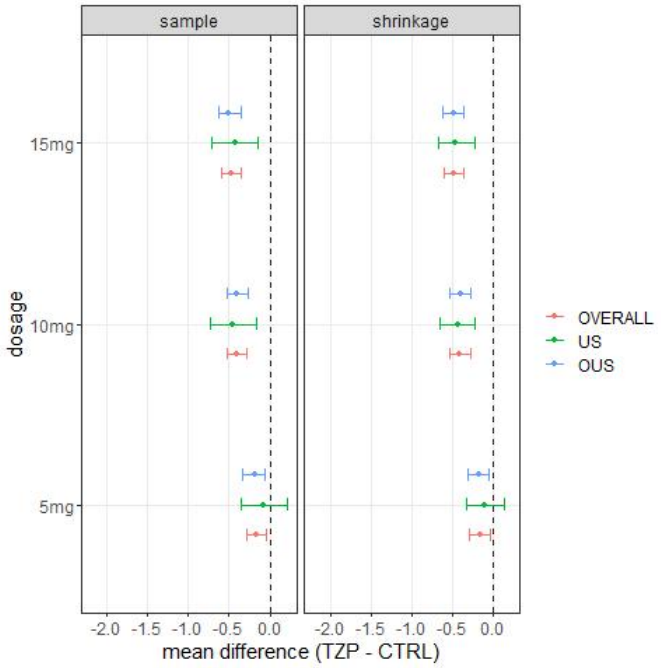


Figure 10: Subgroup Analysis for Region (US or Outside US), SURPASS-2

With respect to the subgroup-by-treatment interaction, significant interaction effects due to difference in effect sizes (though the effects were all found in favor of tirzepatide) were found in:

- AGE groups in SURPASS-1 (p-value[13] = 0.02), SURPASS-3 (p-value = 0.0003), and SURPASS-4 (p-value = 0.01)
- RACE groups in SURPASS-2 (p-value = 0.02)
- SEX groups in SURPASS-4 (p-value = 0.025)

In SURPASS-2, the significant race-by-subgroup interaction may be triggered by the imbalanced sample sizes across the different racial profiles. In particular, the limited sample sizes for the Asian group and for the Black group resulted in estimates of high variability. This issue was mitigated by the shrinkage method which leveraged information from the overall estimate.

In SURPASS-1, -3, and 4, a higher dose of TZP was linked to a more prominent treatment effect in the young groups (< 65 years). Nonetheless, in the old groups (≥ 65 years), this clear dose-response relationship was either absent (SURPASS-3 and -4), or reversed (SURPASS-1). For SURPASS-1, the age-by-treatment interaction may be due to the imbalanced and limited sample sizes of the two age groups, and the shrinkage method was able to partially bridge the gap between the sample estimates of the two groups. On the other hand, with decent sample sizes for both age groups in SURPASS-3 and -4, it was not clear what trigged significant interactions between age and subgroups.

[13] All the p-values for testing interaction effects were two-sided.

Reference ID: 4938321

4.2 Effect Modification of Baseline HbA1c

In this section, the impact of baseline HbA1c (%) on the primary efficacy outcome was explored. Specifically, for each study, a simple linear model that regressed the primary efficacy endpoint on baseline A1c (%) was constructed for each treatment arm based on completers data. The slopes of the regression models were presented in Table 26 below.

Table 26: Slopes (S.E.) of Linear Regressions (A1c change from baseline ~ baseline A1c) Base on Completers Dataset

	SURPASS-1	SURPASS-2	SURPASS-3	SURPASS-4	SURPASS-5
Comparator	-0.53 (0.15)	-0.70 (0.04)	-0.66 (0.05)	-0.63 (0.04)	-0.40 (0.11)
TZP 5mg	-0.72 (0.09)	-0.76 (0.04)	-0.71 (0.06)	-0.71 (0.06)	-0.90 (0.07)
TZP 10mg	-1.07 (0.11)	-0.71 (0.04)	-0.68 (0.05)	-0.71 (0.06)	-0.84 (0.07)
TZP 15mg	-0.85 (0.09)	-0.77 (0.04)	-0.80 (0.05)	-0.68 (0.05)	-0.87 (0.10)

In all treatment arms of each study, the amount of A1c reduction from baseline increased with higher baseline A1c levels. The rate of increasing, however, differed across treatment arms, with most noticeable difference observed between the placebo arms and the TZP arms in the two placebo-controlled trials: SURPASS-1 and SURPASS-5. Specifically, for both studies, the slope for the placebo arm was flatter than the TZP arms, as demonstrated in Table 26, and visualized in Figures 11 and 12.

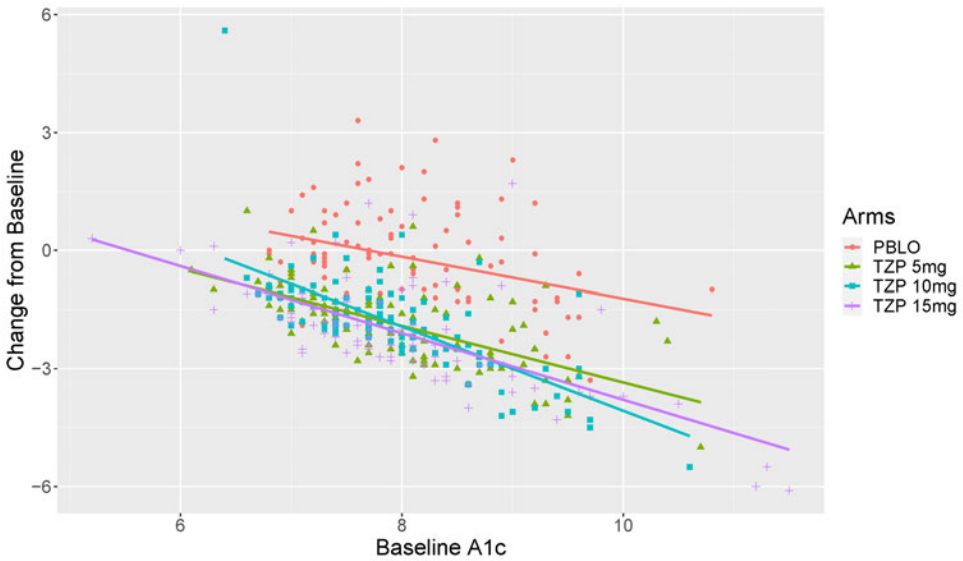


Figure 11: A1c Change from Baseline (%) vs Baseline A1c (%), Completers Data, SURPASS-1[14]

[14] The regression models for PBLO, TZP 5mg, TZP 10 mg and TZP 15 mg: y = -0.53x + 4.07, y = -0.72x + 3.85, y = -1.07x + 6.66, and y = -0.85x + 4.69, respectively, where x=A1c baseline.

Reference ID: 4938321

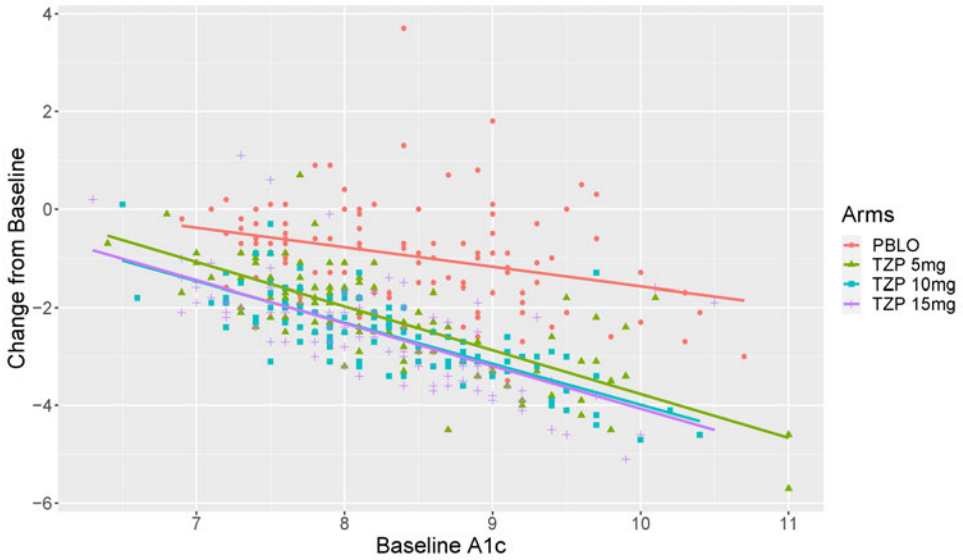


Figure 12: A1c Change from Baseline (%) vs Baseline A1c (%), Completers Data, SURPASS-5[15]

To understand the implications of the differences in slopes, consider a comparison between the placebo (slope = -0.4%) and the TZP 5mg (slope = -0.9%) in SURPASS-5. The difference in these slopes was 0.50%, which implies that with every 1% increase in A1c baseline, the difference in treatment effect increased by 0.5% between these two arms. For further illustration, consider a baseline A1c value of 8%. Based on the regression models, the average change from baseline was -0.78% for the placebo arm, and -2.01% for the TZP 5mg arm, leading to a difference of 1.23%. Now, consider a baseline A1c of 9%. For this subject, the average change from baseline was -1.18% for the placebo arm, and -2.91% for the TZP 5mg arm, resulting in a difference of 1.73%. Thus, for every 1% increase in HbA1c baseline, the difference in treatment effect (as measured by A1c change from baseline) increased by 0.5% (1.73% – 1.23%).

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

No major statistical issues were identified in this review. Significant interaction effects were detected between age and treatment in SURPASS-1, -3 and -4, and between race and treatment in SURPASS-2. These interactions were caused by differences in effect sizes instead of effect directions and were mitigated by shrinkage analyses.

5.2 Collective Evidence

Five Phase III pivotal trials have been conducted to evaluate the use of tirzepatide (as a monotherapy or in combination with other diabetic medications) under three doses: 5mg, 10mg

[15] The regression models for PBLO, TZP 5mg, TZP 10 mg and TZP 15 mg are y = -0.40x + 2.42, y = -0.90x + 5.19, y = -0.84 x+ 4.41, and y = -0.87x + 4.66, respectively, where x=A1c baseline.

Reference ID: 4938321

and 15mg. The efficacy of tirzepatide was compared with placebo, insulin glargine, insulin degludec, and semaglutide 1mg. The rate for missing primary endpoint measures ranged from 4.2% to 9.6% across the five studies. Missing primary endpoints were imputed based on data collected from retrieved dropouts, or from the placebo arm if insufficient retrieved dropouts. Statistical analyses for the primary and key secondary endpoints were performed using the methods pre-specified in the SAPs and agreed upon by the Agency.

Based on the results from the efficacy analyses (Tables 12, 14, 15 and 16), all three doses of tirzepatide achieved statistically significant superiority with large treatment effect size on the primary efficacy endpoint as well as on all the key secondary endpoints, in comparison to both placebo and active controls (with the exception of TZP 5mg vs semaglutide (SURPASS-2) on incidence of HbA1c < 7%). Results from the sensitivity analyses exhibited consistent efficacy results as the primary efficacy analyses (Table 13). Subgroup analyses on the primary efficacy endpoint suggested that the efficacy of tirzepatide were homogeneous across subgroups by age (< 65 years or ≥ 65 years), gender (Male or Female), race (Asian, Black, or White, etc.) and region (the US or outside the US) (Tables 21 through 25). Analyses on the safety database did not find an increased risk of Level 2 or Level 3 hypoglycemic events among the tirzepatide-treated subjects compared to subjects treated with placebo, semaglutide, insulin degludec, or insulin glargine (Tables 17 through 20).

5.3 Conclusions and Recommendations

The statistical findings have shown robust and superior effectiveness of tirzepatide 5mg, 10mg and 15mg in treating adults with T2DM, when compared to placebo, semaglutide 1mg, insulin glargine and insulin degludec, and applied as a monotherapy or in combination with other anti-diabetic medications. These results were able to provide adequate statistical evidence to support the proposed indication for T2DM. The statistical reviewer recommends an approval of the proposed indication.

5.4 Labeling Recommendations

The labelling for tirzepatide is still under review. Major statistical revisions for Section 14 include the following:

(b) (4)

Other edits include:

- Under the efficacy result table, add a footnote that specifies the analysis method for the binary endpoint: incidence of HbA1c < 7%
- List out all the major races (and their percentages) included in the studies (b) (4)
- Add the definition for “inadvertent enrollment”
- For clarity, replace the phrase (b) (4) with “placebo-based multiple imputation”.

Reference ID: 4938321

- Explicitly list out all the covariates for the ANOCVA of the primary analysis (b) (4)
- Specify the types of p-values (one-sided or two sided) used for the efficacy analyses.

Reference ID: 4938321

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

WENDA TU
02/15/2022 01:15:27 PM

YOONHEE KIM
02/15/2022 01:18:35 PM

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Science
Office of Biostatistics

Statistical Review and Evaluation

CARCINOGENICITY STUDY

IND/NDA Number:	NDA 215866
Drug Name:	Tirzepatide (LY3298176)
Indication(s):	For use as an adjunct to diet and exercise to improve glycemic control in adults with type 2 diabetes mellitus (T2DM).
Studies	One Two Year Subcutaneous Injection Carcinogenicity Study in Rats and One Six-Month Subcutaneous Injection Carcinogenicity Study in rasH2 Transgenic Mice.
Applicant:	Sponsor: Eli Lilly And Co. Lilly Corporate Center, Indianapolis, Indiana 46285, Usa (b) (4)
Documents Reviewed:	Electronic submission, dated: September 15, 2021 via SN0001 Electronic data submitted on November 23rd, 2021 via SN0018.
Review Priority:	Standard
Biometrics Division:	Division of Biometrics -VI
Statistical Reviewer:	Malick Mbodj, Ph.D.
Concurring Reviewer:	Karl Lin, Ph.D.
Medical Division:	Division of Diabetes, Lipid Disorders, and Obesity (DDLO)
Reviewing Pharmacologist:	Elena Braithwaite, PhD
Project Manager:	Lindsey Kelly, Pharm.D.
Keywords:	Carcinogenicity, Dose response

Table of Contents

1.Background....2

2.Rat Study....3
2.1. Sponsor's analyses....3
2.1.1. Survival analysis....3
Sponsor's findings....4
2.1.2. Tumor data analysis....4
Adjustment for the multiplicity:....4
Sponsor's findings....5
2.2 Reviewer's analyses....5
2.2.1 Survival analysis....5
Reviewer's findings:....5
2.2.2. Tumor data analysis....5
Multiple testing adjustments:....6
Reviewer's findings:....6

3.Mouse Study....7
3.1. Sponsor's analyses....8
3.1.1 Survival analysis....8
Sponsor's findings....8
3.1.2 Tumor data analysis....8
Sponsor's findings....9
3.2 Reviewer's analyses....9
3.2.1 Survival analysis....9
Reviewer's findings....9
3.2.2 Tumor data analysis....9
Multiple testing adjustment:....9
Reviewer's findings:....9

4.Summary....10
Rat Study:....10
Mouse Study:....11

5.Appendix....13

6.References:....29

1. Background

In this submission, the sponsor included reports of two animal carcinogenicity studies, one in regular rats and one in transgenic mice. These studies were intended to assess the potential carcinogenicity of Tirzepatide

(LY3298176), when administered twice weekly by subcutaneous injection at appropriate drug levels for about 104 weeks in rats and 26 weeks in mice. Results of this review have been discussed with the reviewing pharmacologist Dr. Braithwaite.

In this review, the phrase "dose response relationship" (trend) refers to the linear component of the effect of treatment, and not necessarily to a strictly increasing or decreasing mortality or tumor incidence rate as dose increases.

2. Rat Study

In this study two separate experiments were conducted, one in male rats and one in female rats. In each of these two experiments there were three treated groups and one vehicle control group. Two hundred and forty Crl:CD(SD) rats of each sex were assigned to three treated groups and one vehicle control group by a stratified randomization scheme designed to achieve similar group mean body weights in equal size of 60 animals, as indicated in Table 1. The dose levels for the three treated groups were 0.15, 0.50, and 1.5 mg/kg for both male and female rats, for up to 104 weeks. In this review, these dose groups were referred to as the low, medium, and high dose group, respectively. Rats were dosed via subcutaneous injection twice weekly (i.e. the first and fifth days of each week) for up to 88 weeks (males) or 84 weeks (females) at a volume of 1 mL/kg, the vehicle control group received the vehicle [10 mM Tris, 150 mM sodium chloride (NaCl), and 0.02% Polysorbate 80, pH 7.0 ± 0.2.], administered by subcutaneous injection for about 104 weeks in the same manner as the treated groups.

Table 1: Experimental Design in Rat Study

Group Name	Group N0.	Dose Level (mg/kg/day)		Number of Animal	
		Male	Female	Males	Females
Vehicle Control	1	0	0	60	60
Low	2	0.15	0.15	60	60
Medium	3	0.50	0.50	60	60
High	4	1.5	1.5	60	60

During the study period all animals were observed for general health/mortality and moribundity twice daily (a.m. and p.m.), abnormal findings were recorded throughout the study. Cage side observations were conducted for each carcinogenicity animal once daily during the dosing phase, except on days when detailed observations were conducted. Abnormal findings or an indication of normal was recorded. In addition, for each grossly visible or palpable mass, the time of onset, location, size, appearance, and progression was recorded. Any animal showing signs of severe debility or intoxication, and if determined to suffer excessively was euthanized. Observations will include, but will not be limited to, evaluation for reaction to treatment. Histopathological examinations were performed on all animals found dead or killed moribund or sacrificed at the end of the experiment. Body weights were recorded once during the predose phase, before dosing on Day 1 and 5 of the dosing phase, twice weekly thereafter to Week 26 of the dosing phase, on Day 183 of the dosing phase, then weekly (based on Day 183 of the dosing phase) thereafter to Week 88 (males) or 84 (females) of the dosing phase.

2.1. Sponsor's analyses

2.1.1. Survival analysis

In the sponsor’s analysis, the tests for survival comparisons were performed with a two-sided risk for increasing and decreasing mortality with dose. Tests were performed for dose response and for each test

article-treated group against the control using Kaplan-Meier product-limit estimates, along with log-rank and Wilcoxon tests. These were performed using the LIFETEST procedure in Statistical Analysis Software (SAS). The time to death or sacrifice (in weeks, calculation detailed in the following) was the dependent variable. The test article-treated groups were included as the strata. Animals with a death or sacrifice status recorded as a scheduled sacrifice (interim or terminal) or accidental death will be censored in the analysis.

Sponsor's findings:

Sponsor's analysis showed the numbers of rats surviving to their terminal necropsy were 19 (32%), 30 (50%), 27 (45%), and 20 (33%), in the vehicle control, low, medium, and high dose groups, in male rats, respectively, and 20 (33%), 30 (50%), 28 (47%), and 37 (62%), in the vehicle control, low, medium, and high dose groups, in female rats, respectively. The sponsor's report showed statistically significant trend of a decrease in mortality across the vehicle control group and the treated groups for female rats (p=0.0052 and p=0.0046, using log-rank and Wilcoxon test, respectively). Also, the sponsor's report concluded that, there was a statistically significant decrease in mortality in the low and high dose groups, when compare to the vehicle control group, in female rats (P = 0.0476 and P = 0.0008 for the Log-Rank test, respectively and P = 0.0492 and P = 0.0007 for the Wilcoxon test, respectively)

2.1.2. Tumor data analysis

In the sponsor's analysis, tests to compare tumor incidence were performed, with a one-sided risk for increasing incidence with dose. Tests were performed for dose response and for each test article-treated group against the control group. Nonpalpable, nonfatal, and fatal tumors were analyzed by the IARC asymptotic fixed interval-based prevalence test (Peto et al., 1980) and death rate tests, respectively. The cut-off points for the interval-based test were Weeks 0 to 52, 53 to 73, 74 to before terminal sacrifice, and the terminal sacrifice, for males, and Weeks 0 to 52, 53 to 69, 70 to before terminal sacrifice, and the terminal sacrifice, for females. Actual dose levels were used as the scores. Fatal and non-fatal tumors were analyzed together, with separate stratum for each using the death-rate method and the prevalence methods, respectively. Tumors of uncertain context of observation were included in the analysis as non-fatal. The test was implemented using PROC MULTTEST in the SAS system (SAS, 2008). In the case of sparse tables (<10 total tumor bearing animals in the groups analyzed for the trend or pairwise test), the exact form of the test was used. Otherwise, the asymptotic version of the test was used. Observable or palpable (superficial as in mammary or skin) tumors were analyzed using the methods previously described for analyzing survival, using the time to death or time of detection of the tumor (in weeks) as a surrogate for the tumor onset time. For each given tumor type, statistical analysis was performed if the incidence in at least one dosed group was increased by at least two occurrences over either of the control groups.

Site or tumor combinations were also statistically analyzed. The criteria for combination were based on Guidelines for combining neoplasms for evaluation of rodent carcinogenicity studies (McConnell et al., 1986) and as indicated by the Study Pathologist. Incidences of multiple-organ and combined neoplastic findings, such as hemangioma, fibrosarcoma, and endometrial stromal polyp, were counted by animal, not tissue type.

Adjustment for the multiplicity:

Adjustment followed the recommendations made in the current FDA guidelines (Food and Drug Administration, 2001). The Study Pathologist determined whether a tumor type was rare or common. ·

Sponsor's findings:

Following the multiple testing adjustment method described above, the sponsor's report showed statistically significant increasing trend for c-cell adenomas, and the combined c-cell adenomas and carcinoma in the thyroid, through the high, mid and low dose in both male and female rats. The pairwise comparisons showed statistically significant increases in the low, medium and high dose group for the incidences of c-cell adenomas, and the combined c-cell adenomas and carcinoma in the thyroid, when compared to the vehicle control group in males and females. Also, the pairwise comparisons showed a statistically significant increase in the medium dose group for the incidences of c-cell carcinoma in the thyroid, when compared to the vehicle control group in males.

2.2 Reviewer's analyses

To verify sponsor's analysis and to perform additional analyses suggested by the reviewing pharmacologist, this reviewer independently performed the survival and tumor data analyses. Data used in this reviewer's analyses were provided by the sponsor electronically on November 23rd, 2021 via SN0018.

2.2.1 Survival analysis

In the reviewer's analysis, intercurrent mortality data were analyzed using the Kaplan-Meier product limit method. The Kaplan-Meier's curves were presented graphically for male and female rats separately. The dose response relationship and homogeneity of survival distributions were tested for the treatment groups using the Likelihood Ratio test and the Log-Rank test. The intercurrent mortality data are given in Tables 1A and 1B in the appendix for male and female rats, respectively. The Kaplan-Meier curves for survival rate are given in Figures 1A and 1B in the appendix for male and female rats, respectively. Results of the tests for dose response relationship and homogeneity of survivals, are given in Tables 2A and 2B in the appendix for male and female rats, respectively.

Reviewer's findings:

This reviewer's analysis showed the numbers of rats surviving to their terminal necropsy were 19 (32%), 30 (50%), 27 (45%), and 20 (33%), in the vehicle control, low, medium, and high dose groups, in male rats, respectively, and 20 (33%), 30 (50%), 28 (47%), and 37 (62%), in the vehicle control, low, medium, and high dose groups, in female rats, respectively. This reviewer's analysis showed statistically significant dose response relationship in the mortality of female rats (P=0.0042). The pairwise comparisons also showed statistically significant decrease in mortality between the high dose group, and the vehicle control group in female rats (p=0.0009). Also, the log-rank test showed statistically significant difference in mortality between the low dose group, and the vehicle control group in female rats (p=0.0476).

2.2.2. Tumor data analysis

In the reviewer's analysis, the tumor data were analyzed for dose response relationship across vehicle control group and the treated groups, as well as the pairwise comparisons of vehicle control group with each of the treated groups using the Poly-k method described in the paper of Bailer and Portier (1988) and

Bieler and Williams (1993). In this method, an animal that lives the full study period ($w_{\max}$) or dies before the terminal sacrifice with development of the tumor type being tested gets a score of s_h =1. An animal that dies at Week w_h without development of the given tumor type before the end of the study gets a score of $s_h = \left(\frac{w_h}{w_{\max}} \right)^k$ <1. The adjusted group size is defined as Σs_h . As an interpretation, an animal with score s_h =1 can be considered as a whole animal, while an animal with score s_h <1 can be considered as a partial animal. The adjusted group size Σs_h is equal to N (the original group size) if all animals live up to the end of the study or if each animal develops the given tumor being tested, otherwise the adjusted group size is less than N. These adjusted group sizes are then used for the dose response relationship (or the pairwise comparison) tests using the Cochran-Armitage test. One critical point for Poly-k test is the choice of the appropriate value of k. For long term 104-week standard rat and mouse studies, a value of k=3 is suggested in the literature [Gebregziabher and Hoel (2009), Moon et al. (2003), Portier, et al. (1986)]. Hence, this reviewer used k=3 for the analysis of the data. Based on the intent to treat (ITT) principle Wmax was considered as 105 for both male and female rats.

For the calculation of p-values, if there were less than 10 tumor bearing animals across all treatment groups for a given tumor type, the exact tests based on the discrete permutation distribution were used, with (0, 15, 50, and 150 for both male and female rats) as scores, and asymptotic tests were used for tumor types with higher incidences. The tumor rates and the p-values of the tested tumor types are listed in Tables 3A and 3B in the appendix for male rats and female rats, respectively.

Multiple testing adjustments:

Following the FDA draft guidance for the carcinogenicity study design and data analysis 2001, for the two-year rat study this reviewer used significance levels of 0.005 and 0.025 for common and rare tumors, respectively in dose response relationship (trend) tests and significance levels of 0.01 and 0.05 for common and rare tumors, respectively in pairwise comparisons.

A tumor is defined as a rare tumor if the published spontaneous rate or the spontaneous rate of the vehicle control of the tumor is less than 1%, and a common tumor is defined as one with tumor rate greater than or equal to 1%.

Reviewer's findings:

Table 2: Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship or the pairwise Comparisons
Treated Groups and Control Group in Rats

Sex	**Organ Name**	**Tumor Name**	**0 mg Vehicle Cont (N=60) P - Trend**	**0.15 mg Low (N=60) P - VC vs. L**	**0.5 mg Med (N=60) P - VC vs. M**	**1.5 mg High (N=60) P - VC vs. H**
Male	Thyroid	B-Adenoma, C-cell	6/60 (27) <0.0001*	19/60 (37) 0.0169@	29/60 (41) 0.0001*	38/60 (46) <0.0001*
		M-Carcinoma, C-cell	0/60 (24) 0.4799	4/60 (29) 0.0811	8/60 (30) 0.0056*	2/60 (26) 0.2653
		B-Adenoma/ M-Carcinoma, C-cell	6/60 (27) <0.0001*	20/60 (37) 0.0099*	33/60 (43) <0.0001*	39/60 (46) <0.0001*
Female	Thyroid	B-Adenoma, C-cell	4/59 (24) 0.0006*	19/60 (35) 0.0035*	30/60 (40) <0.0001*	31/60 (43) <0.0001*

Sex	Organ Name	Tumor Name	0 mg Vehicle Cont (N=60) P - Trend	0.15 mg Low (N=60) P - VC vs. L	0.5 mg Med (N=60) P - VC vs. M	1.5 mg High (N=60) P - VC vs. H
		M-Carcinoma, C-cell	1/59 (22) 0.0561	3/60 (26) 0.3708	2/60 (25) 0.5489	6/60 (30) 0.1129
		Adenoma/ Carcinoma, c-cell	5/59 (24) 0.0004*	20/60 (35) 0.0055*	30/60 (40) <0.0001*	34/60 (45) <0.0001*

& X/ZZ (YY): X=number of tumor bearing animals; YY=mortality weighted total number of animals; ZZ=unweighted total number of animals observed;
*: Statistically significant at 0.005 and 0.025 level for common and rare tumor or 0.01 and 0.05 level for common and rare tumors for tests of dose response relationship and pairwise comparison, respectively.
@: Not statistically significant at 0.01 for common tumors for pairwise comparison.

Following the multiple testing adjustment method described above, this reviewer's analysis showed a statistically significant dose response relationship in tumor incidences with increased LY3298176 dose across the vehicle control and the treated groups for the incidence of benign adenoma, c-cell and the combined benign adenomas and malignant carcinoma, c-cell, in the thyroid, for both male and female rats (p-value <0.0001 and < 0.0001, and < 0.0006 and < 0.0004, respectively). The pairwise comparisons showed a statistically significant increase in the medium and high dose group for the incidences of benign adenoma, c-cell and the combined benign adenomas and malignant carcinoma, c-cell, in the thyroid, in male rats, when compare to the vehicle control group (p-value < 0.0001 and < 0.0001, < 0.0001 and < 0.0001, respectively). Also, the incidences in the thyroid, of malignant carcinoma, c-cell in the medium dose group, and the combined benign adenomas and malignant carcinoma, c-cell in the low dose group were statistically significant higher, when compared to the vehicle control group (p-value =0.0056 and =0.0099, respectively), in male rats. However, in Female rats, the pairwise comparisons showed a statistically significant increase in the low, medium and high dose group for the incidences of benign adenoma, c-cell and the combined benign adenomas and malignant carcinoma, c-cell, in the thyroid, when compare to the vehicle control group (p-value = 0.0035 and < 0.0001 and < 0.0001, and =0.0055 and < 0.0001 and < 0.0001, respectively).

3. Mouse Study

Two separate experiments were conducted, one in male mice and one in female mice. In each of these two experiments there were three treated groups, one vehicle control group, and one positive control group. One hundred and ten hemizygous transgenic RasH2 mice of each sex were assigned randomly to one of the five groups by a stratified randomization scheme designed to achieve similar group mean body weights in equal size of 25 animals except the positive control group which had 10 animals. The dose levels for the three treated groups were 1, 3, and 10 mg/kg for both male and female mice, for up to 26 weeks, as indicated in Table 3. In this review, these dose groups would be referred to as the low, medium, and high dose group, respectively. Animals were dosed via subcutaneous injection in the dorsal region (Groups 1 to 4 The positive control article is N-methyl-N-nitrosourea (MNU), also known as N-nitroso-N-methylurea, by intraperitoneal (IP) injection on Day 1 at a dose of 75 mg/kg. This group was included to verify sensitivity of the test system to detect carcinogenicity effect. The vehicle control groups received daily oral vehicle control article only [10 mM Tris, 150 mM sodium chloride (NaCl), and 0.02% polysorbate 80, pH 7.0 ± 0.2.], were administered in the same manner as the treated groups.

Table 3: Experimental Design in Mouse Study

Group Name	Group N0.	Dose Level (mg/kg/day) Male	Female	Number of Animal Males	Females
Vehicle Control	1	0	0	25	25
Low	2	1	1	25	25
Medium	3	3	3	25	25
High	4	10	10	25	25
Positive control	5	75 NMU	75 NMU	10	10

The positive control was administered with 1 intraperitoneal (i.p.) injections of urethane in saline on Days 1.

During the study period, all animals were observed for general health/mortality, signs of pain or distress, twice daily (a.m. and p.m.), abnormal findings were recorded throughout the study. Clinical sign observations were performed once during the predose phase and for each carcinogenicity animal prior to dosing on Day 1, weekly (based on Day 1) thereafter to Week 26, and on Day 182 of the dosing phase. In addition, for each grossly visible or palpable mass, the time of onset, location, size, appearance, and progression was recorded. Any animal showing signs of severe debility or intoxication, and if determined to suffer excessively was euthanized. Observations will include, but will not be limited to, evaluation for reaction to treatment. Histopathological examinations were performed on all animals found dead or killed moribund or sacrificed at the end of the experiment. Body weights were measured once during the predose phase, before dosing on Days 1 and 5 of the dosing phase, twice weekly (based on Days 1 and 5) thereafter, and on Day 182 of the dosing phase

3.1. Sponsor's analyses

3.1.1 Survival analysis

The sponsor used similar methodologies to analyze the mouse survival data as those used to analyze the rat survival data.

Sponsor’s findings:

Sponsor’s analysis showed the numbers of mice surviving to their terminal necropsy were 24 (96%), 24 (96%), 24 (96%), and 25 (100%), in vehicle control group, low, medium, and high dose groups in male mice, respectively, and 23 (92%), 24 (96%), 24 (96%), and 23 (92%), in female mice, respectively. The sponsor’s report concluded that there were no statistically significant findings in survival rate in either sex of mice.

.

3.1.2 Tumor data analysis

The sponsor used similar methodologies to analyze the mouse tumor data as those used to analyze the rat tumor data. The cut off points for the interval-based test were Weeks 0 to 13, 14 to before terminal sacrifice, and the terminal sacrifice.

Multiple testing adjustment:

As there were no statistically significant findings, indication of a possible treatment effect did not need to be assessed on the basis of rare or common tumor type, in line with the current FDA guidelines (Food and Drug Administration Draft Guidance for Industry, 2001).

Sponsor's findings:

the sponsor's report concluded that, there were no statistically significant ($p<0.05$) increasing or trends in tumor incidence through the high dose group in either male or female mice when the positive control group was excluded. Also, the sponsor's report concluded that, there were no statistically significant difference in tumor incidence between the high dose group and the vehicle control group in either male or female mice

3.2 Reviewer's analyses

Similar to the rat study, this reviewer independently performed the survival and tumor data analyses of the mouse study. For the analysis of the survival data and the tumor data of the mouse study, this reviewer used similar methodologies that were used for the analyses of the survival and tumor data of the rat study. Data used in this reviewer's analyses were provided by the sponsor electronically.

3.2.1 Survival analysis

The intercurrent mortality data are given in Tables 4A and 4B in the appendix for male and female mice, respectively. The Kaplan-Meier curves for death rate are given in Figures 2A and 2B in the appendix for male and female mice, respectively. Results for test of dose response relationship and homogeneity of survivals among treatment groups are given in Tables 5A and 5B in the appendix for male and female mice, respectively.

Reviewer's findings:

This reviewer's analysis showed the numbers of mice surviving to their terminal necropsy were 24 (96%), 24 (96%), 24 (96%), and 25 (100%), in vehicle control group, low, medium, and high dose groups in male mice, respectively, and 23 (92%), 24 (96%), 24 (96%), and 23 (92%), in female mice, respectively. This reviewer's analysis showed no statistically significant increase or decrease in mortality across the vehicle control group and the three treated groups in either sex of mice. The pairwise comparisons showed no statistically significant increase or decrease in mortality between the treated groups, and the vehicle control group in either sex of mice.

For both males and females, the positive control group had a statistically significant increase ($p<0.001$) in unscheduled death compared with the vehicle control group.

3.2.2 Tumor data analysis

The tumor rates and the p-values of the tumor types tested for dose response relationship and the pairwise comparisons of the vehicle controls and treated groups are given in Table 6A and 6B in the appendix for male and female mice, respectively.

Multiple testing adjustment:

For multiplicity testing adjustment in a transgenic mouse study, this reviewer used the level of significance of 0.05 for tests of positive dose responses and for pairwise increase comparisons for both common and rare tumor types.

Reviewer's findings:

Following the multiple testing adjustment method described above, this reviewer's analysis showed no tumor types with a statistically significant dose response relationship in tumor incidences with increased LY3298176 dose. The pairwise comparisons also showed no tumor types with a statistically significant increase in tumor incidences in LY3298176 treated groups, when compare to the vehicle control group in either male or female mice.

Also, this reviewer's analyses showed Statistically significant in positive control group for the incidences of malignant lymphoma in the hematolymphoid system, benign papilloma, squamous cell, in the skin/subcutis, and benign papilloma, squamous cell, in the stomach, nonglandular, compared to the vehicle control group in both male and female mice (P<0.0001, <0.0001, <0.0001, and <0.0001, =0.0031, <0.0055, respectively). Also, the incidence of malignant papilloma, squamous cell, in the skin/subcutis, were statistically higher, when compared to the vehicle control group in male mice (p-value =0.0246).

4. Summary

In this submission, the sponsor included reports of two animal carcinogenicity studies, one in regular rats and one in transgenic mice. These studies were intended to assess the potential carcinogenicity of Tirzepatide (LY3298176), when administered twice weekly by subcutaneous injection at appropriate drug levels for about 104 weeks in rats and 26 weeks in mice.

Rat Study:

In this study two separate experiments were conducted, one in male rats and one in female rats. In each of these two experiments there were three treated groups and one vehicle control group. Two hundred and forty Crl:CD(SD) rats of each sex were assigned to three treated groups and one vehicle control group by a stratified randomization scheme designed to achieve similar group mean body weights in equal size of 60 animals, as indicated in Table 1. The dose levels for the three treated groups were 0.15, 0.50, and 1.5 mg/kg for both male and female rats, for up to 104 weeks. In this review, these dose groups were referred to as the low, medium, and high dose group, respectively. Rats were dosed via subcutaneous injection twice weekly (i.e. the first and fifth days of each week) for up to 88 weeks (males) or 84 weeks (females) at a volume of 1 mL/kg, the vehicle control group received the vehicle [10 mM Tris, 150 mM sodium chloride (NaCl), and 0.02% Polysorbate 80, pH 7.0 ± 0.2.], administered by subcutaneous injection for about 104 weeks in the same manner as the treated groups.

This reviewer's analysis showed the numbers of rats surviving to their terminal necropsy were 19 (32%), 30 (50%), 27 (45%), and 20 (33%), in the vehicle control, low, medium, and high dose groups, in male rats, respectively, and 20 (33%), 30 (50%), 28 (47%), and 37 (62%), in the vehicle control, low, medium, and high dose groups, in female rats, respectively. This reviewer's analysis showed statistically significant dose response relationship in the mortality of female rats (P=0.0042). The pairwise comparisons also showed statistically significant decrease in mortality between the high dose group, and the vehicle control group in female rats (p=0.0009). Also, the log-rank test showed statistically significant difference in mortality between the low dose group, and the vehicle control group in female rats (p=0.0476).

For tumor data, following the multiple testing adjustment method described above, this reviewer's analysis showed a statistically significant dose response relationship in tumor incidences with increased LY3298176 dose across the vehicle control and the treated groups for the incidence of benign adenoma, c-cell and the combined benign adenomas and malignant carcinoma, c-cell, in the thyroid, for both male and female rats (p-value <0.0001 and < 0.0001, and < 0.0006 and < 0.0004, respectively). The pairwise comparisons showed a statistically significant increase in the medium and high dose group for the incidences of benign adenoma, c-

cell and the combined benign adenomas and malignant carcinoma, c-cell, in the thyroid, in male rats, when compare to the vehicle control group (p-value < 0.0001 and < 0.0001, < 0.0001 and < 0.0001, respectively). Also, the incidences in the thyroid, of malignant carcinoma, c-cell in the medium dose group, and the combined benign adenomas and malignant carcinoma, c-cell in the low dose group were statistically significant higher, when compared to the vehicle control group (p-value =0.0056 and =0.0099, respectively), in male rats. However, in Female rats, the pairwise comparisons showed a statistically significant increase in the low, medium and high dose group for the incidences of benign adenoma, c-cell and the combined benign adenomas and malignant carcinoma, c-cell, in the thyroid, when compare to the vehicle control group (p-value = 0.0035 and < 0.0001 and < 0.0001, and =0.0055 and < 0.0001 and < 0.0001, respectively)..

Mouse Study:

Two separate experiments were conducted, one in male mice and one in female mice. In each of these two experiments there were three treated groups, one vehicle control group, and one positive control group. One hundred and ten hemizygous transgenic RasH2 mice of each sex were assigned randomly to one of the five groups by a stratified randomization scheme designed to achieve similar group mean body weights in equal size of 25 animals except the positive control group which had 10 animals. The dose levels for the three treated groups were 1, 3, and 10 mg/kg for both male and female mice, for up to 26 weeks, as indicated in Table 3. In this review, these dose groups would be referred to as the low, medium, and high dose group, respectively. Animals were dosed via subcutaneous injection in the dorsal region (Groups 1 to 4 The positive control article is N-methyl-N-nitrosourea (MNU), also known as N-nitroso-N-methylurea, by intraperitoneal (IP) injection on Day 1 at a dose of 75 mg/kg. This group was included to verify sensitivity of the test system to detect carcinogenicity effect. The vehicle control groups received daily oral vehicle control article only [10 mM Tris, 150 mM sodium chloride (NaCl), and 0.02% polysorbate 80, pH 7.0 ± 0.2.], were administered in the same manner as the treated groups.

This reviewer's analysis showed the numbers of mice surviving to their terminal necropsy were 24 (96%), 24 (96%), 24 (96%), and 25 (100%), in vehicle control group, low, medium, and high dose groups in male mice, respectively, and 23 (92%), 24 (96%), 24 (96%), and 23 (92%), in female mice, respectively. This reviewer's analysis showed no statistically significant increase or decrease in mortality across the vehicle control group and the three treated groups in either sex of mice. The pairwise comparisons showed no statistically significant increase or decrease in mortality between the treated groups, and the vehicle control group in either sex of mice.

For both males and females, the positive control group had a statistically significant increase (p<0.001) in unscheduled death compared with the vehicle control group.

For tumor data, following the multiple testing adjustment method described above, this reviewer's analysis showed no tumor types with a statistically significant dose response relationship in tumor incidences with increased LY3298176 dose. The pairwise comparisons also showed no tumor types with a statistically significant increase in tumor incidences in LY3298176 treated groups, when compare to the vehicle control group in either male or female mice.

Also, this reviewer's analyses showed Statistically significant in positive controls group for the incidences of malignant lymphoma in the hematolymphoid system, benign papilloma, squamous cell, in the skin/subcutis, and benign papilloma, squamous cell, in the stomach, nonglandular, compared to the vehicle control groups in both male and female mice (P<0.0001, <0.0001, <0.0001, and <0.0001, =0.0031, <0.0055, respectively). Also, the incidence of malignant papilloma, squamous cell, in the skin/subcutis, were statistically higher, when compared to the vehicle control group in male mice (p-value =0.0246).

Malick Mbodj, Ph.D.
Mathematical Statistician

Concur: Karl Lin, Ph.D. Team Leader, DBVI

cc:
Archival NDA 215866 - LY3298176
Dr. Tsong Dr. Lin
Lindsey Kelly Dr. Braithwaite
Dr. Rahman Dr. Collins

5. Appendix

Table1A: Intercurrent Mortality Rate
Male Rats

	0 mg|kg|day		0.15 mg|kg|day		0.50 mg|kg|day		1.50 mg|kg|day	
Week	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**
0 - 52	9	15.00	5	8.33	6	10.00	5	8.33
53 - 73	13	36.67	14	31.67	12	30.00	19	40.00
74 - 88	19	68.33	11	50.00	15	55.00	16	66.67
Ter. Sac.	19	31.67	30	50.00	27	45.00	20	33.33
Total	60	100.00	60	100.00	60	100.00	60	100.00

Animals were assigned to the terminal sacrifice strata based on the sacrifice status recorded

Table1B: Intercurrent Mortality Rate
Female Rats

	0 mg|kg|day		15 mg|kg|day		50 mg|kg|day		150 mg|kg|day	
Week	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**
0 - 52	4	6.67	3	5.00	5	8.33	3	5.00
53 - 73	23	45.00	16	31.67	16	35.00	9	20.00
74 - 84	13	66.67	11	50.00	11	53.33	11	38.33
Ter. Sac.	20	33.33	30	50.00	28	46.67	37	61.67
Total	60	100.00	60	100.00	60	100.00	60	100.00

Animals were assigned to the terminal sacrifice strata based on the sacrifice status recorded

Table 2A: Intercurrent Mortality Comparison for
Male Rats

Test Statistics	P-value for Vehicle Cont. Low, Med, high	P-value for Vehicle Cont. vs Low	P-value for Vehicle Cont. vs Med	P-value for Vehicle Cont. vs High
Dose-Response (Likelihood Ratio)	0.4663	0.0553	0.1553	0.8187
Homogeneity (Log-Rank)	0.1511	0.0531	0.1506	0.8166

Table 2B: Intercurrent Mortality Comparison for
Female Rats

Test Statistics	P-value for Vehicle Cont. Low, Med, high	P-value for Vehicle Cont. vs Low	P-value for Vehicle Cont. vs Med	P-value for Vehicle Cont. vs High
Dose-Response (Likelihood Ratio)	0.0042*	0.0508	0.1354	0.0009*
Homogeneity (Log-Rank)	0.0090*	0.0476*	0.1297	0.0008*

* = statistically significant at the 0.05 significance level

Table3A: Tumor Rates and P-Values for Dose Response Relationship and the pairwise comparisons

Male Rats Poly-3 Test

Organ Name	Tumor Name	0 mg Vehicle Cont (N=60) P - Trend	15 mg Low (N=60) P - C vs. L	50 mg Med (N=60) P - C vs. M	150 mg High (N=60) P - C vs. H
Adipose, Other	M-Liposarcoma	0/60 (24) 0.2427	0/60 (27) NC	0/60 (27) NC	1/60 (25) 0.5102
	M-Schwannoma, malignant	1/60 (25) 1.0000	0/60 (27) 1.0000	0/60 (27) 1.0000	0/60 (25) 1.0000
Adrenal, Cortex	B-Adenoma	0/60 (24) 0.3127	1/60 (28) 0.5385	0/60 (27) NC	1/60 (25) 0.5102
Adrenal, Medulla	B-Pheochromocytoma	2/60 (25) 0.9281	1/60 (28) 0.9018	1/60 (28) 0.9018	0/60 (25) 1.0000
	M-Malignant pheochromocytoma	2/60 (25) 1.0000	0/60 (27) 1.0000	0/60 (27) 1.0000	0/60 (25) 1.0000
Brain	M-Malignant astrocytoma	1/60 (25) 0.9451	1/60 (28) 0.7823	0/60 (27) 1.0000	0/60 (25) 1.0000
Cecum	B-Lipoma	0/60 (24) 0.2427	0/60 (27) NC	0/60 (27) NC	1/60 (25) 0.5102
Eye	M-Melanoma, amelanotic	1/59 (25) 0.4357	0/60 (27) 1.0000	0/60 (27) 1.0000	1/60 (26) 0.7647
Gingiva	M-Osteosarcoma	1/60 (25) 1.0000	0/60 (27) 1.0000	0/60 (27) 1.0000	0/60 (25) 1.0000
Hematolymphoid System	M-Histiocytic sarcoma	1/60 (25) 0.6840	0/60 (27) 1.0000	2/60 (28) 0.5432	0/60 (25) 1.0000
	M-Lymphoma	0/60 (24) 0.2500	0/60 (27) NC	0/60 (27) NC	1/60 (26) 0.5200
Kidney	M-Liposarcoma	0/60 (24) 0.1856	0/60 (27) NC	1/60 (27) 0.5294	1/60 (25) 0.5102
Liver	B-Adenoma, hepatocellular	1/60 (25) 1.0000	0/60 (27) 1.0000	0/60 (27) 1.0000	0/60 (25) 1.0000
Lung	M-Malignant mesothelioma	0/60 (24) 0.7692	1/60 (28) 0.5385	0/60 (27) NC	0/60 (25) NC
Lymph Node, Mesenteric	B-Hemangioma	1/59 (25) 0.9429	2/60 (28) 0.5432	0/60 (27) 1.0000	0/60 (25) 1.0000
	M-Hemangiosarcoma	1/59 (25) 1.0000	0/60 (27) 1.0000	0/60 (27) 1.0000	0/60 (25) 1.0000
Mammary Gland	B-Fibroadenoma	2/40 (16) 0.9452	1/45 (22) 0.9336	1/49 (23) 0.9387	0/48 (19) 1.0000
Pancreas	B-Adenoma, islet cell	1/60 (25) 0.6013	2/60 (28) 0.5432	1/60 (27) 0.7738	1/60 (25) 0.7551

Male Rats Poly-3 Test

Organ Name	Tumor Name	0 mg Vehicle Cont (N=60) P - Trend	15 mg Low (N=60) P - C vs. L	50 mg Med (N=60) P - C vs. M	150 mg High (N=60) P - C vs. H
	M-Carcinoma, islet cell	1/60 (25) 1.0000	0/60 (27) 1.0000	0/60 (27) 1.0000	0/60 (25) 1.0000
Parathyroid	B-Adenoma	2/56 (23) 1.0000	0/56 (26) 1.0000	0/55 (24) 1.0000	0/53 (22) 1.0000
Pituitary	B-Adenoma, pars distalis	34/60 (46) 0.4190	37/60 (48) 0.4530	37/58 (48) 0.4530	37/60 (48) 0.4530
	B-Adenoma, pars intermedia	0/60 (24) 0.2451	0/60 (27) NC	0/58 (26) NC	1/60 (25) 0.5102
	M-Schwannoma, malignant	0/60 (24) 0.2524	0/60 (27) NC	0/58 (26) NC	1/60 (26) 0.5200
Skin/Subcutis	B-Adenoma, sebaceous	1/60 (25) 1.0000	0/60 (27) 1.0000	0/60 (27) 1.0000	0/60 (25) 1.0000
	B-Fibroma	3/60 (26) 0.5193	1/60 (28) 0.9527	2/60 (28) 0.8468	2/60 (26) 0.8251
	B-Lipoma	1/60 (25) 0.9451	1/60 (28) 0.7823	0/60 (27) 1.0000	0/60 (25) 1.0000
	B-Papilloma, squamous cell	0/60 (24) 0.8271	3/60 (29) 0.1560	1/60 (27) 0.5294	0/60 (25) NC
	M-Fibrosarcoma	0/60 (24) 0.3643	1/60 (28) 0.5385	3/60 (28) 0.1482	1/60 (25) 0.5102
	B-Fibroma / M-Fibrosarcoma	3/60 (26) 0.4026	2/60 (29) 0.8565	5/60 (29) 0.4172	3/60 (26) 0.6660
	M-Malignant melanoma	1/60 (25) 1.0000	0/60 (27) 1.0000	0/60 (27) 1.0000	0/60 (25) 1.0000
	M-Malignant schwannoma	1/60 (25) 1.0000	0/60 (27) 1.0000	0/60 (27) 1.0000	0/60 (25) 1.0000
Spinal Cord	M-Malignant astrocytoma	0/60 (24) 0.2524	0/60 (27) NC	0/59 (26) NC	1/60 (26) 0.5200
Stomach, Nonglandular	B-Papilloma, squamous cell	1/60 (25) 1.0000	0/60 (27) 1.0000	0/60 (27) 1.0000	0/60 (25) 1.0000
Subcutaneous Injection Site, B	B-Adenoma, sebaceous cell	0/60 (24) 0.7692	1/60 (28) 0.5385	0/60 (27) NC	0/60 (25) NC
	B-Fibroma	0/60 (24) 0.7692	1/60 (28) 0.5385	0/60 (27) NC	0/60 (25) NC
Subcutaneous Injection Site, D	B-Fibroma	1/60 (25) 1.0000	0/60 (27) 1.0000	0/60 (27) 1.0000	0/60 (25) 1.0000
Testis	B-Interstitial cell tumor	1/58 (25) 0.3910	0/60 (27) 1.0000	3/60 (28) 0.3496	1/60 (25) 0.7551

Male Rats Poly-3 Test

Organ Name	Tumor Name	0 mg Vehicle Cont (N=60) P - Trend	15 mg Low (N=60) P - C vs. L	50 mg Med (N=60) P - C vs. M	150 mg High (N=60) P - C vs. H
Thymus	B-Thymoma	0/56 (23) 0.0658	0/52 (23) NC	0/55 (25) NC	2/58 (25) 0.2660
	M-Schwannoma, malignant	0/56 (23) 0.7604	1/52 (24) 0.5106	0/55 (25) NC	0/58 (24) NC
Thyroid	B-Adenoma, C-cell	6/60 (27) <0.0001*	19/60 (37) 0.0169	29/60 (41) 0.0001*	38/60 (46) <0.0001*
	B-Adenoma, follicular cell	1/60 (25) 0.8318	1/60 (28) 0.7823	1/60 (27) 0.7738	0/60 (25) 1.0000
	M-Carcinoma, C-cell	0/60 (24) 0.4799	4/60 (29) 0.0811	8/60 (30) 0.0056*	2/60 (26) 0.2653
	B-Adenoma, C-cell / M-Carcinoma, C-cell	6/60 (27) <0.0001*	20/60 (37) 0.0099*	33/60 (43) <0.0001*	39/60 (46) <0.0001*
Zymbal Gland	M-Carcinoma	1/56 (23) 0.9396	1/53 (25) 0.7757	0/54 (24) 1.0000	0/51 (20) 1.0000

Table3B: Tumor Rates and P-Values for Dose Response Relationship and the pairwise comparisons

Female Rats Poly-3 Test

Organ Name	Tumor Name	0 mg Vehicle Cont (N=60) P - Trend	15 mg Low (N=60) P - C vs. L	50 mg Med (N=60) P - C vs. M	150 mg High (N=60) P - C vs. H
Adipose, Other	B-Hibernoma	0/60 (21) 0.3682	1/60 (25) 0.5435	0/60 (24) NC	1/60 (28) 0.5714
Adrenal, Cortex	B-Adenoma	1/60 (22) 0.3226	1/60 (25) 0.7863	1/60 (24) 0.7768	2/60 (28) 0.5914
Cervix	M-Sarcoma, stromal	0/60 (21) 0.2857	0/60 (25) NC	0/60 (24) NC	1/60 (28) 0.5714
	M-Schwannoma, malignant	0/60 (21) 0.5258	0/60 (25) NC	1/60 (24) 0.5333	0/60 (27) NC
Hematolymphoid System	M-Histiocytic sarcoma	0/60 (21) 0.6718	1/60 (25) 0.5435	1/60 (25) 0.5435	0/60 (27) NC
	M-Lymphoma	1/60 (22) 1.0000	0/60 (25) 1.0000	0/60 (24) 1.0000	0/60 (27) 1.0000
Kidney	B-Adenoma, tubule cell	0/60 (21) 0.2784	0/60 (25) NC	0/60 (24) NC	1/60 (27) 0.5625
	M-Liposarcoma	0/60 (21) 0.7835	1/60 (25) 0.5435	0/60 (24) NC	0/60 (27) NC
Liver	B-Adenoma, hepatocellular	0/59 (21) 0.2784	0/60 (25) NC	0/60 (24) NC	1/60 (27) 0.5625
Mammary Gland	B-Adenoma	1/58 (22) 0.2082	0/60 (25) 1.0000	1/59 (24) 0.7768	2/60 (28) 0.5914
	B-Fibroadenoma	19/58 (32) 0.9775	21/60 (38) 0.7216	19/59 (35) 0.7473	12/60 (33) 0.9828
	B-Fibroma	0/58 (21) 0.7835	1/60 (25) 0.5435	0/59 (24) NC	0/60 (27) NC
	M-Carcinoma	17/58 (32) 0.9890	8/60 (29) 0.9894	11/59 (31) 0.9522	6/60 (31) 0.9990
	B-Adenoma/ B-Fibroadenoma/ M-Carcinoma	32/58 (41) 0.9965	28/60 (42) 0.9202	28/59 (41) 0.8939	17/60 (36) 0.9989
Mesentery	M-Mesothelioma, malignant	0/60 (21) 0.2784	0/60 (25) NC	0/60 (24) NC	1/60 (27) 0.5625
Ovary	B-Granulosa/theca cell tumor	0/60 (21) 0.7813	1/60 (25) 0.5435	0/59 (23) NC	0/60 (27) NC
	B-Luteoma	0/60 (21) 0.5258	0/60 (25) NC	1/59 (24) 0.5333	0/60 (27) NC
Pancreas	M-Mesothelioma, malignant	0/59 (21) 0.2784	0/60 (25) NC	0/60 (24) NC	1/60 (27) 0.5625

Female Rats Poly-3 Test

Organ Name	Tumor Name	0 mg Vehicle Cont (N=60) P - Trend	15 mg Low (N=60) P - C vs. L	50 mg Med (N=60) P - C vs. M	150 mg High (N=60) P - C vs. H
Parathyroid	B-Adenoma	0/56 (20) 0.5319	0/59 (24) NC	1/58 (24) 0.5455	0/57 (26) NC
Pituitary	B-Adenoma, pars distalis	44/60 (49) 0.9821	37/59 (45) 0.9137	33/60 (43) 0.9764	32/60 (45) 0.9953
	M-Carcinoma	2/60 (23) 0.3318	0/59 (24) 1.0000	0/60 (24) 1.0000	2/60 (28) 0.7661
	B-Adenoma, pars distalis/ M-Carcinoma	46/60 (50) 0.9753	37/59 (45) 0.9601	33/60 (43) 0.9910	34/60 (46) 0.9965
Skin/Subcutis	B-Basal cell tumor	0/60 (21) 0.2857	0/60 (25) NC	0/60 (24) NC	1/60 (28) 0.5714
	B-Papilloma, squamous cell	2/60 (23) 1.0000	0/60 (25) 1.0000	0/60 (24) 1.0000	0/60 (27) 1.0000
Spleen	M-Mesothelioma, malignant	0/60 (21) 0.2784	0/60 (25) NC	0/60 (24) NC	1/60 (27) 0.5625
Stomach, Glandular	M-Mesothelioma, malignant	0/60 (21) 0.2784	0/60 (25) NC	0/60 (24) NC	1/60 (27) 0.5625
Subcutaneous Injection Site, B	M-Sarcoma	1/60 (22) 1.0000	0/60 (25) 1.0000	0/60 (24) 1.0000	0/60 (27) 1.0000
Subcutaneous Injection Site, C	M-Fibrosarcoma	0/60 (21) 0.5306	0/60 (25) NC	1/60 (25) 0.5435	0/60 (27) NC
Thymus	B-Thymoma	0/55 (19) 0.3620	1/55 (23) 0.5476	0/59 (23) NC	1/54 (25) 0.5682
	M-Malignant thymoma	0/55 (19) 0.7889	1/55 (23) 0.5476	0/59 (23) NC	0/54 (25) NC
Thyroid	B-Adenoma, C-cell	4/59 (24) 0.0006*	19/60 (35) 0.0035*	30/60 (40) <0.0001*	31/60 (43) <0.0001*
	B-Adenoma, follicular cell	1/59 (22) 1.0000	0/60 (25) 1.0000	0/60 (24) 1.0000	0/60 (27) 1.0000
	M-Carcinoma, C-cell	1/59 (22) 0.0561	3/60 (26) 0.3708	2/60 (25) 0.5489	6/60 (30) 0.1129
	B-Adenoma, C-cell/ M-Carcinoma, C-cell	5/59 (24) 0.0004*	20/60 (35) 0.0055*	30/60 (40) <0.0001*	34/60 (45) <0.0001*
	M-Carcinoma, follicular cell	0/59 (21) 0.2784	0/60 (25) NC	0/60 (24) NC	1/60 (27) 0.5625
Uterus	B-Polyp, endometrial stromal	3/60 (23) 0.2820	1/60 (25) 0.9545	4/60 (26) 0.5713	4/60 (29) 0.6323

Female Rats Poly-3 Test

Organ Name	Tumor Name	0 mg Vehicle Cont (N=60) P - Trend	15 mg Low (N=60) P - C vs. L	50 mg Med (N=60) P - C vs. M	150 mg High (N=60) P - C vs. H
	M-Carcinoma, squamous cell	0/60 (21) 0.3595	1/60 (25) 0.5435	0/60 (24) NC	1/60 (27) 0.5625
Vagina	B-Tumor, granular cell, ben*	0/60 (21) 0.5258	0/60 (25) NC	1/60 (24) 0.5333	0/60 (27) NC
	M-Carcinoma, squamous cell	0/60 (21) 0.5258	0/60 (25) NC	1/60 (24) 0.5333	0/60 (27) NC
	M-Sarcoma, stromal	0/60 (21) 0.7835	1/60 (25) 0.5435	0/60 (24) NC	0/60 (27) NC
	M-Schwannoma, malignant	1/60 (22) 1.0000	0/60 (25) 1.0000	0/60 (24) 1.0000	0/60 (27) 1.0000

& X/ZZ (YY): X=number of tumor bearing animals; YY=mortality weighted total number of animals; ZZ=unweighted total number of animals observed;
NC = Not calculable
*: Statistically significant at 0.025 for rare tumor in dose response relationship

Figure 1A: Kaplan-Meier Survival Curves for
Male Rats

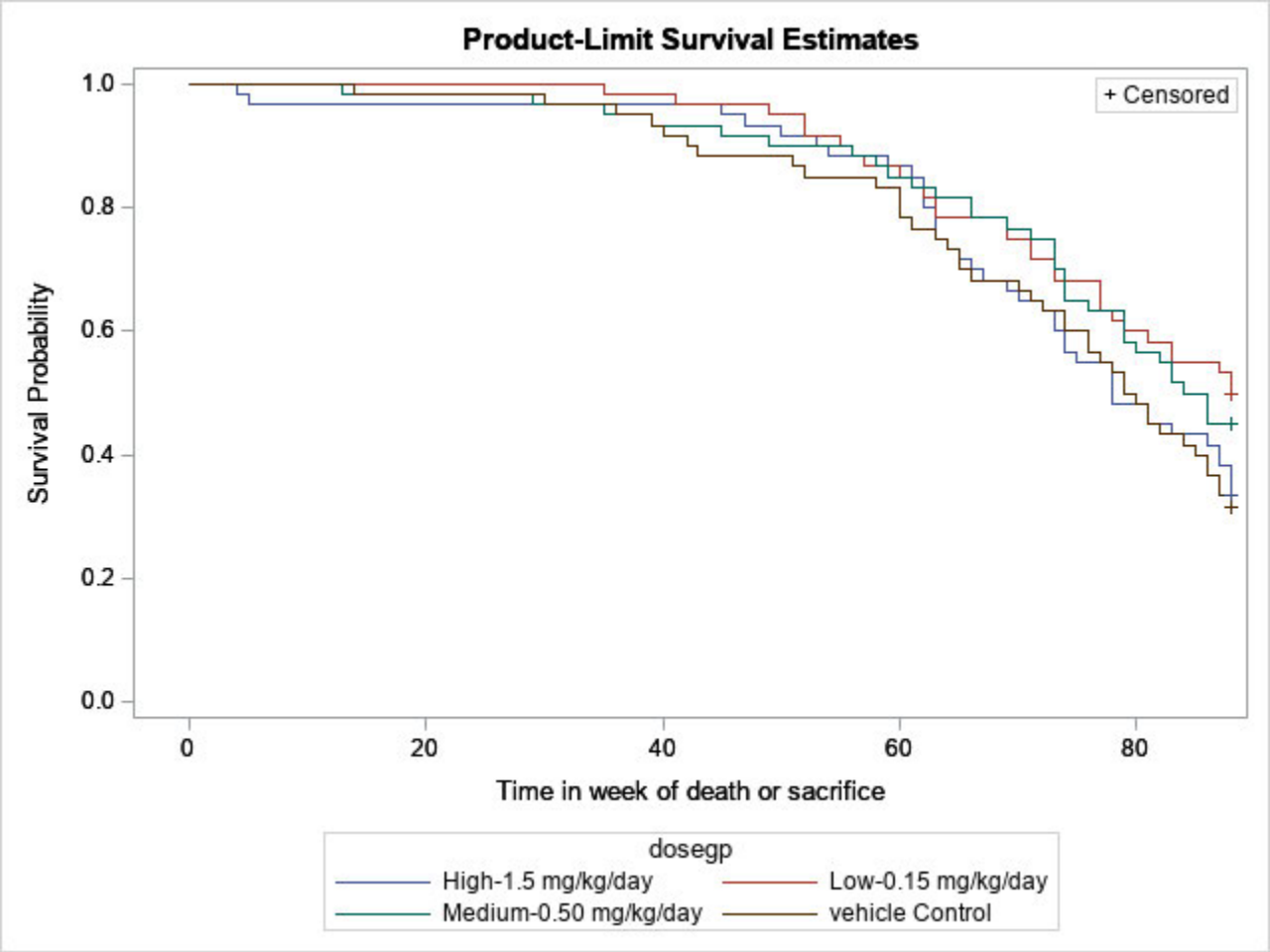


Figure 1B: Kaplan-Meier Survival Curves **for Female Rats**

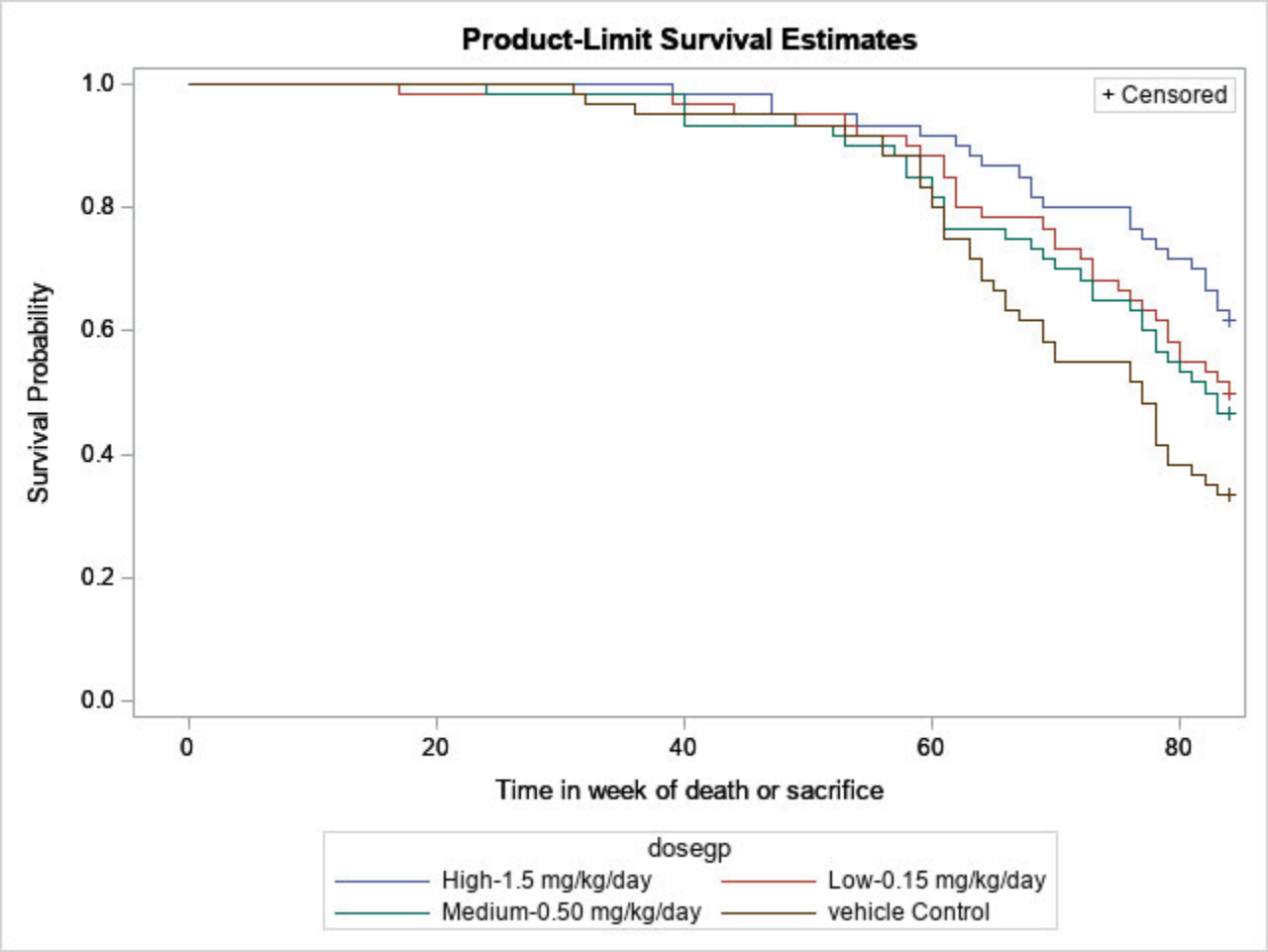


Table4A: Intercurrent Mortality Rate
Male Mice

	0 mg|kg|day Vehicle Control		1 mg|kg|day Low		3 mg|kg|day Medium		10 mg|kg|day High		75 mg|kg|day Positive	
Week	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**
0 - 13	.	.	.	.	.	.	.	.	1	10.00
14 - 26	1	4.00	1	4.00	1	4.00	.	.	8	90.00
Ter. Sac.	24	96.00	24	96.00	24	96.00	25	100.00	1	10.00
Total	25	100.00	25	100.00	25	100.00	25	100.00	10	100.00

Animals were assigned to the terminal sacrifice strata based on the sacrifice status recorded

Table4B: Intercurrent Mortality Rate
Female Mice

	0 mg|kg|day Vehicle Control		1 mg|kg|day Low		3 mg|kg|day Medium		10 mg|kg|day High		75 mg|kg|day Positive	
Week	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**	**No. of Death**	**Cum. %**
0 - 13	2	8.00	.	.	1	4.00	.	.	2	20.00
14 - 26	.	.	1	4.00	.	.	2	8.00	8	100.00
Ter. Sac.	23	92.00	24	96.00	24	96.00	23	92.00	.	.
Total	25	100.00	25	100.00	25	100.00	25	100.00	10	100.00

Animals were assigned to the terminal sacrifice strata based on the sacrifice status recorded

Table 5A: Intercurrent Mortality Comparison for
Male Mice

Test Statistics	P-value for Vehicle Control Low, Med, high	P-value for Vehicle Control vs Low	P-value for Vehicle Control vs Med	P-value for Vehicle Control vs High	P-value for Vehicle Control vs Positive Control
Dose-Response (Likelihood Ratio)	0.2596	0.9885	0.9885	0.2390	<.0001*
Homogeneity (Log-Rank)	0.7978	0.9885	0.9885	0.3173	<.0001*

*: Statistically significant at 0.025 for rare tumor in dose response relationship

Table 5B: Intercurrent Mortality Comparison for
Female Mice

Test Statistics	P-value for Vehicle Control Low, Med, high	P-value for Vehicle Control vs Low	P-value for Vehicle Control vs Med	P-value for Vehicle Control vs High	P-value for Vehicle Control vs Positive Control
Dose-Response (Likelihood Ratio)	0.7761	0.5362	0.5362	0.9667	<.0001*
Homogeneity (Log-Rank)	0.8687	0.5396	0.5396	0.9667	<.0001*

*: Statistically significant at 0.025 for rare tumor in dose response relationship

Table 6A: Tumor Rates and P-Values for Dose Response Relationship and the pairwise Comparisons

Male Mice Using Poly-3 test

Organ Name	Tumor Name	0 mg Vehicle Cont (N=25) P - Trend	1 mg Low (N=25) P - VC vs. L	3 mg Med(N=25) P - VC vs. M	10 mg High (N=25) P - VC vs. H	75 mg Posi (N=10) P - VC vs. PC
Eye	M-Hemangiosarcoma	0/25 (24) 0.7576	1/25 (25) 0.5102	0/25 (25) NC	0/25 (25) NC	0/10 (5) NC
Hematolymphoid System	M-Lymphoma	0/25 (24) NC	0/25 (25) NC	0/25 (25) NC	0/25 (25) NC	9/10 (9) <0.0001*
Kidney	B-Hemangioma	0/25 (24) 0.3195	1/25 (25) 0.5102	0/25 (25) NC	1/25 (25) 0.5102	0/10 (5) NC
Liver	B-Adenoma, Hepatocellular	1/25 (24) 1.0000	0/25 (25) 1.0000	0/25 (25) 1.0000	0/25 (25) 1.0000	0/10 (5) 1.0000
Lung	B-Adenoma, Bronchiolo-Alveolar	2/25 (24) 0.9837	2/25 (25) 0.7110	0/25 (25) 1.0000	0/25 (25) 1.0000	0/10 (5) 1.0000
	M-Carcinoma, Bronchiolo-Alveolar	0/25 (24) 0.5051	0/25 (25) NC	2/25 (25) 0.2551	0/25 (25) NC	0/10 (5) NC
	B-Adenoma/M-Carcinoma, Bronchiolo-Alveolar	2/25 (24) 0.9298	2/25 (25) 0.7110	2/25 (25) 0.7110	0/25 (25) 1.0000	0/10 (5) 1.0000
Skin/Subcutis	B-Papilloma, Squamous Cell	0/25 (24) 0.2525	0/25 (25) NC	0/25 (25) NC	1/25 (25) 0.5102	7/10 (8) <0.0001*
	M-Carcinoma, Squamous Cell	0/25 (24) NC	0/25 (25) NC	0/25 (25) NC	0/25 (25) NC	2/10 (5) 0.0246*
Spleen	M-Hemangiosarcoma	3/25 (24) 0.9961	3/25 (25) 0.6864	0/25 (25) 1.0000	0/25 (25) 1.0000	0/10 (5) 1.0000
Stomach, Nonglandular	B-Papilloma, Squamous Cell	0/25 (24) NC	0/25 (25) NC	0/25 (25) NC	0/25 (25) NC	7/10 (9) <0.0001*
	M-Carcinoma, Squamous Cell	0/25 (24) NC	0/25 (25) NC	0/25 (25) NC	0/25 (25) NC	1/10 (5) 0.1724
Thymus	M-Mesothelioma	1/25 (25) 1.0000	0/24 (24) 1.0000	0/23 (23) 1.0000	0/24 (24) 1.0000	0/10 (5) 1.0000
Eye/Kidney/Spleen	M-Hemangiosarcoma	3/25 (24) 0.9470	5/25 (25) 0.3748	0/25 (25) 1.0000	1/25 (25) 0.9498	0/10 (5) 1.0000

& X/ZZ (YY): X=number of tumor bearing animals; YY=mortality weighted total number of animals; ZZ=unweighted total number of animals observed;
NC = Not calculable
*: Statistically significant at levels of 0.01 and 0.05 for common and rare tumors, respectively in pairwise comparisons.

Table 6B: Tumor Rates and P-Values for Dose Response Relationship and The pairwise comparisons
Female Mice Using Poly-3 test

Organ Name	Tumor Name	0 mg Vehicle Cont (N=25) P - Trend	1 mg Low (N=25) P - VC vs. L	3 mg Med(N=25) P - VC vs. M	10 mg High (N=25) P - VC vs. H	75 mg Posi (N=10) P-VC vs. PC
Harderian Gland	B-Adenoma	0/25 (23) 0.5053	0/25 (24) NC	1/25 (24) 0.5106	0/25 (24) NC	0/10 (4) NC
	M-Adenocarcinoma	0/25 (23) 0.2526	0/25 (24) NC	0/25 (24) NC	1/25 (24) 0.5106	0/10 (4) NC
Hematolymphoid System	M-Lymphoma	0/25 (23) NC	0/25 (24) NC	0/25 (24) NC	0/25 (24) NC	7/10 (8) <0.0001*
Lung	B-Adenoma, Bronchiolo-Alveolar	2/25 (23) 0.8370	0/25 (24) 1.0000	2/25 (24) 0.7121	0/25 (24) 1.0000	0/10 (4) 1.0000
	M-Carcinoma, Bronchiolo-Alveolar	0/25 (23) NC	0/25 (24) NC	0/25 (24) NC	0/25 (24) NC	1/10 (4) 0.1481
Muscle, Biceps Femoris	M-Hemangiosarcoma	1/25 (24) 1.0000	0/25 (24) 1.0000	0/25 (24) 1.0000	0/25 (24) 1.0000	0/10 (4) 1.0000
Skin/Subcutis	B-Papilloma, Squamous Cell	0/25 (23) NC	0/25 (24) NC	0/25 (24) NC	0/25 (24) NC	3/10 (5) 0.0031*
Spleen	M-Hemangiosarcoma	1/25 (23) 0.6242	2/25 (24) 0.5163	1/25 (24) 0.7660	1/25 (24) 0.7660	0/10 (4) 1.0000
Stomach, Nonglandular	B-Papilloma, Squamous Cell	0/25 (23) NC	0/25 (24) NC	0/25 (24) NC	0/25 (24) NC	3/10 (6) 0.0055*
	M-Carcinoma, Squamous Cell	0/25 (23) NC	0/25 (24) NC	0/25 (24) NC	0/25 (24) NC	1/10 (5) 0.1786
Thymus	B-Thymoma, Benign	0/25 (23) 0.8158	2/25 (24) 0.2553	0/25 (24) NC	0/24 (23) NC	0/9 (4) NC
	M-Mesothelioma	0/25 (23) 0.3198	1/25 (24) 0.5106	0/25 (24) NC	1/24 (24) 0.5106	0/9 (4) NC
Vagina	M-Hemangiosarcoma	1/25 (24) 1.0000	0/25 (24) 1.0000	0/25 (24) 1.0000	0/24 (23) 1.0000	0/10 (4) 1.0000
Muscle/Spleen/ Vagina	M-Hemangiosarcoma	3/25 (25) 0.8394	2/25 (24) 0.8129	1/25 (24) 0.9403	1/25 (24) 0.9403	0/10 (4) 1.0000

& X/ZZ (YY): X=number of tumor bearing animals; YY=mortality weighted total number of animals; ZZ= total number of animals observed;
NC = Not calculable;
*: Statistically significant at levels of 0.01 and 0.05 for common and rare tumors, respectively in pairwise comparisons

.

Figure 2A: Kaplan-Meier Survival Curves for

Male Mice

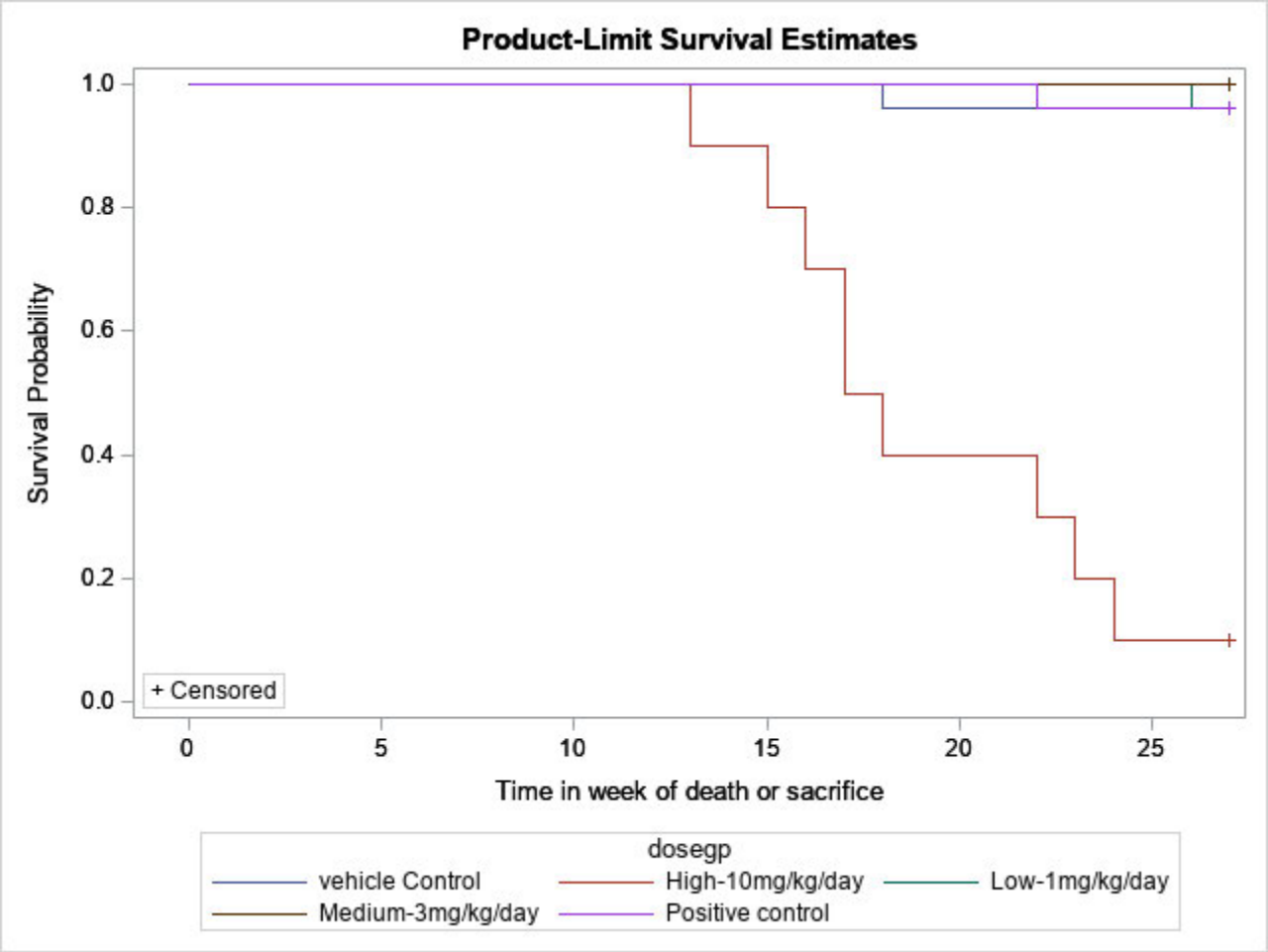
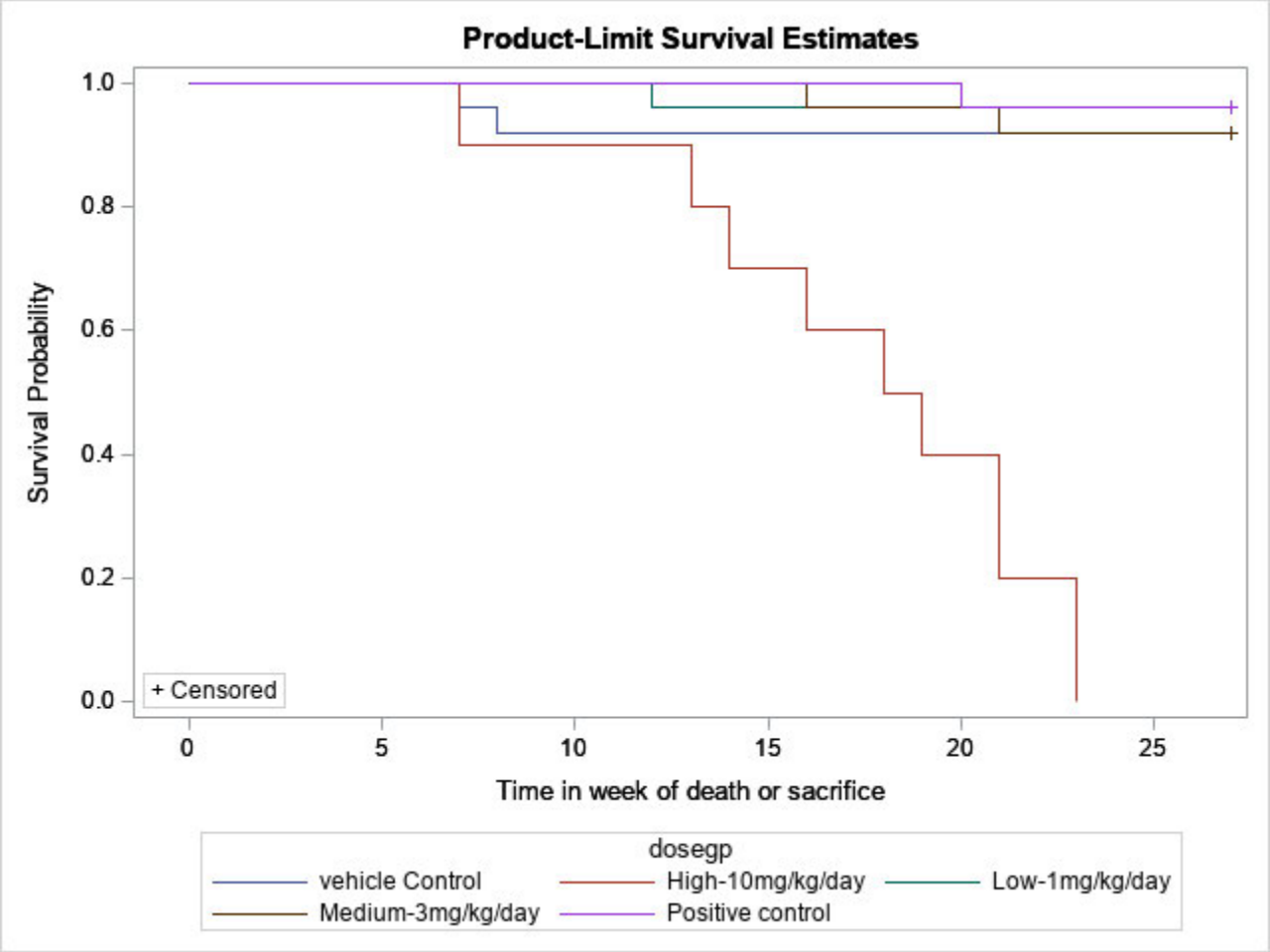


Figure 2B: Kaplan-Meier Survival Curves for
Female Mice



6. References:

1. Bailer AJ, Portier CJ (1988). "Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples." *Biometrics*, 44, 417-431.

2. Bieler, G. S. and Williams, R. L. (1993). "Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity". *Biometrics* 49, 793-801.

3. Gebregziabher M, Hoel D (2009) Applications of the poly-k statistical test to life-time cancer bioassay studies. Hum Ecol Risk Assess 15(5):858{875.

4. Guidance for Industry. Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Statues of Pharmaceuticals (Draft Guidance). U.S. Department of Health and Human Services, Food and Drug Administation, Center for Drug Evaluation and Research (CDER), May 2001.

5. Moon H, Ahn H, Kodell RL, Lee JJ (2003) Estimation of k for the poly-k test with application to animal carcinogenicity studies. Statistics in Medicine 22(16):2619{2636

6. Peto, R., M.C. Pike, N.E. Day, R.G. Gray, P.N. Lee, S. Parish, J. Peto, Richards, and Wahrendorf, "Guidelines for sample sensitive significance test for carcinogenic effects in long-term animal experiments", Long term and short term screening assays for carcinogens: A critical appraisal, International agency for research against cancer monographs, Annex to supplement, World Health Organization, Geneva, 311-426, 1980.

7. Portier C, Hedges J, Hoel D (1986) Age-speci_c models of mortality and tumor onset for historical control animals in the national toxicology programs carcinogenicity experiments. Cancer Research 46:4372{4378

8. Rahman M.A. and Lin K.K, "A comparison of False Positive Rates of Peto and Poly-3 Method For Long Term Carcinogenicity Data Analysis Using Multiple Comparison Adjustment Method Suggested by Lin and Rahman" Journal of Biopharmaceutical Statistics, 18: 949-958, 2008.

9. Tarone RE, "Test for trend in life table analysis", Biometrika 1975, 62: 679-82

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

MALICK MBODJ
02/11/2022 02:13:11 PM

KARL K LIN
02/14/2022 08:29:34 AM
Concur with review.