Horizon Therapeutics U.S.A., Inc
HZNP-TEP-403

A Phase 4, Randomized, Double-masked, Placebo-controlled, Multicenter Trial to Evaluate the Efficacy and Safety of TEPEZZA® in Treating Patients with Chronic (Inactive) Thyroid Eye Disease

02OCT2023

Statistical Analysis Plan

Version 6.0

Prepared by: PPD

Prepared by: PPD Date: ___/___/___

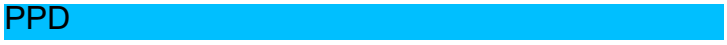
Approved by 

Date: ___/___/___

Horizon Therapeutics U.S.A., Inc.

Approved by: 

Date: ___/___/___



Horizon Therapeutics U.S.A., Inc.

History of Revisions

Effective Date	Version	Author	Description of Changes
25 Jan 2021	1.0	PPD	Original Version
07 May 2021	2.0	PPD	Main reasons for updating the SAP: • Clarify scope of the work for follow-up analysis; • Define open-label ITT/PK analysis set; • Define TEPEZZA baseline and TEPEZZA study day for the open-label treatment period
05 Apr 2022	3.0	PPD	Main reasons for updating the SAP: • Incorporate the changes in protocol amendment version 3.0 and 4.0, which include the increase in sample size and the trial eligibility criteria modification. • Add an additional analysis set (i.e., per protocol set) and remove Open-label ITT, Open-label Safety, and Open-label PK analysis sets (Section 5.3.3) to streamline analysis and summary presentation for easier review • Re-define Analysis Visit Window (Section 5.4.3) • Add the duration of TED to summarize in the demographic and baseline characteristics (Section 7)
09 Jan 2023	4.0	PPD	Main reasons for updating the SAP: • Incorporate the changes in administrative change 1 to Protocol Number: HZN-TEP-403 Version 4.0 (Amendment 3.0), which include re-arranging "Other Efficacy Endpoints" • Update the intercurrent event handling strategy (Section 9.1) • Add sensitivity analyses for the primary efficacy endpoint and key other endpoint (proptosis responder rate at Week 24) (Section 9.2.2 and 9.3.1.2).

Effective Date	Version	Author	Description of Changes
21 Mar 2023	5.0	PPD	Main reasons for updating the SAP: CCI • Mathematical formula for statistical modeling • SAS codes for key efficacy analyses • Summary of restricted medications taken by patients in the trial
02 Oct 2023	6.0	PPD	Main reason for updating the SAP is to add details of Immunogenicity analysis (Section 11)

TABLE OF CONTENTS

LIST OF ABBREVIATIONS **8**

1. INTRODUCTION **10**

2. OBJECTIVES **11**

2.1 Primary Objective 11
2.2 Other Objectives 11
2.3 Pharmacokinetic and Anti-drug Antibody Objectives 11
2.4 Safety and Tolerability Objectives 11
2.5 Exploratory Objective 12

3 INVESTIGATIONAL PLAN **12**

3.1 Overall Study Design and Plan 12
3.2 Study Endpoints 13
3.2.1 Primary Endpoint 13
3.2.2 Other Efficacy Endpoints 13
3.2.3 Pharmacokinetic and Anti-drug Antibody Endpoints 13
3.2.4 Safety and Tolerability Endpoints 14
3.2.5 Exploratory Endpoints 14
3.3 Study Eye 14

4 TYPE OF PLANNED ANALYSIS **14**

4.1 Interim Analysis 14
4.2 Primary Analysis 14
4.3 Follow-up Analysis 14

5 GENERAL STATISTICAL CONSIDERATIONS **15**

5.1 Sample Size 15
5.2 Randomization, Stratification, and Masking 16
5.3 Analysis Set 16
5.3.1 Intent-to-Treat (ITT) 16
5.3.2 Modified Intent-to-Treat (mITT) 16
5.3.3 Per Protocol (PP) 16
5.3.4 Safety 17
5.3.5 PK 17
5.4 Assessment Windows 17
5.5 Missing Data Algorithms 19
5.6 Multiplicity 21

6. PATIENT DISPOSITION **21**

6.1 Disposition 21
6.2 Protocol Deviations 21

7. DEMOGRAPHICS AND BASELINE CHARACTERISTICS **22**

7.1 Substance Use 23
7.2 Medical History 23

7.3 INCLUSION AND EXCLUSION CRITERIA 23

8. TREATMENTS AND MEDICATIONS 23

8.1 PRIOR AND CONCOMITANT MEDICATIONS 23
8.2 CONCOMITANT PROCEDURES 24
8.3 RESTRICTED MEDICATIONS AND SURGICAL PROCEDURES 24
8.4 STUDY TREATMENTS 25
8.4.1 Extent of Exposure 25
8.4.2 Treatment Compliance and Modifications 25

9. EFFICACY ANALYSIS 25

9.1 INTERCURRENT EVENT HANDLING STRATEGY FOR ESTIMANDS 26
9.2 PRIMARY EFFICACY ENDPOINT 26
9.2.1 Primary Analysis Method 26
9.2.2 Sensitivity Analysis Method 27
9.2.3 Supplementary Analysis 29
9.2.4 Subgroup Analysis 29
9.2.5 Exploratory Analysis 29
9.3 OTHER EFFICACY ENDPOINTS 30
9.3.1 Proptosis Responder Rate at Week 24 30
9.3.1.1 Primary Analysis Method 30
9.3.1.2 Sensitivity Analysis Method 31
9.3.2 Change from Baseline at Week 24 in the GO-QoL Questionnaire Appearance and Visual Functioning Subscales 31
9.3.3 CCI 32
9.3.4 CCI 33
9.3.5 CCI 33
9.4 EXPLORATORY ENDPOINTS 34
9.4.1 CCI 34
9.4.2 CCI 34
9.4.3 CCI 34
9.4.4 Clinical Activity Score 34

10. SAFETY ANALYSIS 34

10.1 ADVERSE EVENTS 35
10.1.1 Incidence of Adverse Events 35
10.1.2 Relationship of Adverse Events to Study Drug 36
10.1.3 Severity of Adverse Event 36
10.1.4 Serious Adverse Events 36
10.1.5 Adverse Events Leading to Treatment Discontinuation 36
10.1.6 Adverse Events of Special Interest (AESI) 37
10.2 CLINICAL LABORATORY EVALUATIONS 37
10.3 VITAL SIGN AND WEIGHT MEASUREMENTS 38
10.4 OTHER SAFETY DATA 38

11. IMMUNOGENICITY: ANTI-DRUG ANTIBODY 38

12. PHARMACOKINETICS 39

13. CHANGES IN THE PLANNED ANALYSIS 39

14. REFERENCES 41

15. APPENDICES 42

15.1 SCHEDULE OF ASSESSMENTS 42
15.2 UNIT CONVERSION 46
15.3 GRAVES' OPHTHALMOPATHY QUALITY OF LIFE QUESTIONNAIRE 47
15.4 CCI [redacted] 48
15.5 CCI [redacted] 49
15.5.1 CCI [redacted] 49
15.5.2 CCI [redacted] 50
15.5.3 CCI [redacted] 50
15.6 CCI [redacted] 52
15.6.1 CCI [redacted] 52
15.6.2 CCI [redacted] 52
15.6.3 CCI [redacted] 53

List of Abbreviations

ADA	anti-drug antibody
AE	adverse event
AESI	adverse event of special interest
ATC	Anatomical Therapeutic Chemical
BCVA	best corrected visual acuity
BMI	body mass index
CAS	Clinical Activity Score
CI	confidence interval
eCRF	electronic case report form
FCS	full conditional specification
FDA	Food and Drug Administration
FT3	free triiodothyronine
FT4	free thyroxine
GO-QoL	Graves' Ophthalmopathy Quality of Life
HbA1c	glycated hemoglobin
HLT	high level term
ICE	intercurrent event
IGF	insulin-like growth factor
IGF-1R	insulin-like growth factor-1 receptor
ITT	intent-to-treat
IV	intravenous(ly)
kg	kilogram
LOCF	last observation carried forward
LS	least squares
mAb	monoclonal antibody
MCAR	missing completely at random
MedDRA	Medical Dictionary for Regulatory Activities
mg	milligram
MI	multiple imputation
mITT	modified intent-to-treat
MMRM	Mixed-Model for Repeated-Measures
CCI	[redacted]
NAb	neutralizing antibody
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
PP	Per Protocol
PT	preferred term
PK	pharmacokinetic
PW	premature withdrawal
PW1	premature withdrawal during the double-masked treatment period
PW2	premature withdrawal during the open-label treatment period
Q1	first quartile

Q3	third quartile
q3W	once every 3 weeks
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SOC	system organ class
SMQ	standard MedDRA query
TEAE	treatment-emergent adverse events
TED	thyroid eye disease
TEPEZZA	teprotumumab-trbw; HZN-001
TLF	tables, listings, and figures
TSH	thyroid stimulating hormone
U.S.A./US	United States of America
VAS	visual analog scale
W	week
WHO	World Health Organization

1. Introduction

Thyroid eye disease (TED), also termed Graves' ophthalmopathy/orbitopathy and thyroid-associated ophthalmopathy, is a serious, debilitating and painful autoimmune disease that can, in severe cases, lead to blindness. TED is commonly associated with Graves' hyperthyroidism/disease, but also occurs in a proportion of patients with other autoimmune thyroid diseases, including Hashimoto's thyroiditis.

TEPEZZA ® (teprotumumab-trbw, HZN-001), hereafter referred to as TEPEZZA) is a fully human immunoglobulin G1 monoclonal antibody (mAb) directed against human insulin-like growth factor -1 receptor (IGF-1R). The IGF-1R is a tyrosine kinase cell surface receptor that shares ~50% overall homology with the insulin receptor [Ullrich et al, 1986]. TEPEZZA binds with high affinity and selectivity to the extracellular domain of IGF-1R and prevents its activation by the endogenous ligands, IGF-1 and IGF-2.

TEPEZZA was approved by the US Food and Drug Administration (FDA) on 21 January 2020 for the treatment of TED.

Previous teprotumumab clinical trials for the TED indication include 2 independent, randomized, double-masked, placebo-controlled, parallel-group, multicenter trials (Phase 2 Trial TED01RV [Smith et al, 2017] and Phase 3 Trial HZNP-TEP-301 [Douglas et al, 2020]). In these 2 trials, teprotumumab resulted in statistically significant and clinically relevant improvements in measures that assessed multiple facets of TED (proptosis, inflammation as measured by Clinical Activity Score [CAS], diplopia and quality of life). In addition, the persistence of effect was demonstrated after approximately 1 year off treatment. Consistent results were shown across all efficacy endpoints and all subpopulations.

In the Phase 2 trial, the majority of reported adverse events (AEs) were mild, required no treatment and resolved while patients remained on drug. Hyperglycemia, which was monitored by assessing blood glucose and glycated hemoglobin (HbA1c), was the only mechanism-based anticipated AE clearly identified as related to teprotumumab.

In the Phase 3 trial, teprotumumab was well tolerated and demonstrated an acceptable safety profile, replicating the results of the Phase 2 trial.

This phase 4 study is a randomized, double-masked, placebo-controlled, parallel-group, multicenter trial study. The dose of either TEPEZZA (10 mg/kg for the first infusion and 20 mg/kg for the remaining 7 infusions) or placebo will be administered to patients with chronic (inactive) TED.

This statistical analysis plan (SAP) describes the statistical analysis methods and data presentations to be used in tables, listings, and figures (TLFs) of the primary analysis and follow-up analysis for Study HZNP-TEP-403. The primary analysis is planned after all patients complete the week 24 visit assessments in the double-masked treatment period or withdraw from the study prior to Week 24. The follow-up analysis is planned after all patients complete the 30-day follow-up period or terminate early from the study during the follow-up period. The SAP is

based on protocol: administrative change 1 to Protocol number: HZNP-TEP-403 V4.0 (Amendment 3.0), dated on 15 December 2022 by Horizon Therapeutics U.S.A., Inc.

2. Objectives

The overall objective is to investigate the efficacy, safety and tolerability of TEPEZZA® in comparison to placebo in treating patients with chronic (inactive) TED.

2.1 Primary Objective

The primary objective is to evaluate the effect of TEPEZZA versus placebo on the change of proptosis measurements in the study eye from Baseline at Week 24 in patients with chronic (inactive) TED.

2.2 Other Objectives

1. To evaluate the effect of TEPEZZA versus placebo on the proptosis responder rate (i.e., the percentage of patients with a ≥2-mm reduction from Baseline in the study eye without deterioration [≥2-mm increase] of proptosis in the fellow eye) at Week 24.
2. To evaluate the effect of TEPEZZA versus placebo on the change from Baseline at Week 24 in the Graves' Ophthalmopathy Quality of Life (GO-QoL) questionnaire appearance and visual functioning subscales.
3. To evaluate the effect of TEPEZZA versus placebo on the CCI [redacted]
4. To evaluate the effect of TEPEZZA versus placebo on the CCI [redacted] ≥ [redacted].
5. To evaluate the effect of TEPEZZA versus placebo on the CCI [redacted].

2.3 Pharmacokinetic and Anti-drug Antibody Objectives

1. To evaluate the pharmacokinetic (PK) of TEPEZZA.
2. To evaluate the immunogenicity of TEPEZZA.

2.4 Safety and Tolerability Objectives

To assess safety and tolerability of TEPEZZA versus placebo based on:

1. Adverse events (AEs)
2. AEs of special interest (AESIs) (infusion reactions, hyperglycemia, hearing impairment, new onset inflammatory bowel disease and exacerbation of inflammatory bowel disease)
3. Vital signs
4. Visual acuity
5. Clinical safety laboratory evaluations (hematology, chemistry, thyroid function tests and glycated hemoglobin [HbA1c])

2.5 Exploratory Objective

1. To evaluate the effect of TEPEZZA versus placebo on the CCI [redacted]
2. To evaluate the effect of TEPEZZA versus placebo on the CCI [redacted]
3. To evaluate the effect of TEPEZZA vs placebo on the CCI [redacted]

3 Investigational Plan

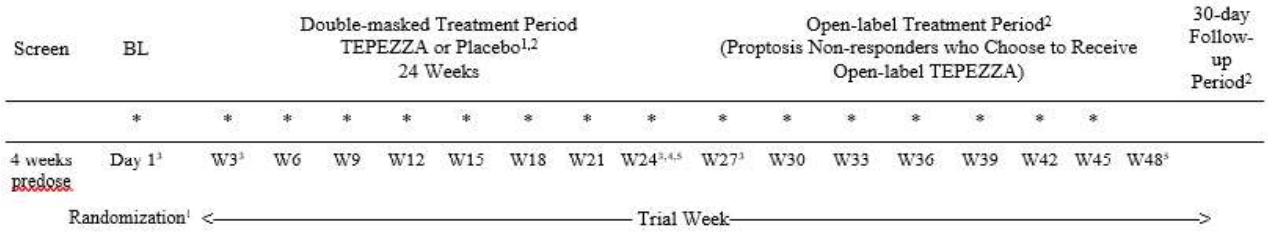
3.1 Overall Study Design and Plan

This is a randomized, double-masked, placebo-controlled, parallel-group, multicenter trial. Patients will be screened for the trial within 4 weeks prior to the Baseline (Day 1) Visit. Approximately 57 patients who meet the trial eligibility criteria will be randomized on Day 1 in a 2:1 ratio to receive 8 infusions of TEPEZZA (10 mg/kg for the first infusion and 20 mg/kg for the remaining 7 infusions) or placebo once every 3 weeks (q3W). All patients will enter a 24-week double-masked treatment period, during which trial drug will be infused on Day 1 (Baseline) and Weeks 3, 6, 9, 12, 15, 18 and 21 (with a final visit at Week 24 of the 24-week treatment period).

At the end of the double-masked treatment period (Week 24), all patients will be assessed for treatment response. The primary treatment response is assessed based on the change in proptosis from Baseline. Proptosis responders were defined as patients with a 2 mm or greater reduction in proptosis from Baseline in the study eye, without deterioration (i.e., 2 mm or great increase) of proptosis in the fellow eye. Proptosis non-responders who have completed the double-masked treatment period may choose to receive 8 infusions of TEPEZZA (10 mg/kg for the first infusion and 20 mg/kg for the remaining 7 infusions) in an open-label fashion q3W at Weeks 24, 27, 30, 33, 36, 39, 42 and 45. These patients will return to the clinic at Week 48 for End-of-Treatment assessments, and patients will be contacted by phone/email 30 days after the Week 48 Visit for safety assessment. Proptosis responders, as well as non-responders who choose not to receive TEPEZZA in an open-label fashion, will enter a 30-day Follow-up Period, during which study drug will not be administered. These patients will be contacted by phone/email 30 days after the Week 24 Visit for safety assessment.

An overview of the trial design is presented in **Figure 1**, and details of trial activities are presented in Section 15.1, Schedule of Assessments.

Figure 1 Schematic of Trial Design



* Infusion of trial drug (TEPEZZA or placebo in double-masked treatment period and open-label TEPEZZA in open-label treatment period).

BL=Baseline; q3W=once every 3 weeks; W=Week.

1. Patients will be randomized in a 2:1 ratio to receive:
   a. TEPEZZA (10 mg/kg on Day 1 followed by 20 mg/kg q3W for the remaining 7 infusions); or
   b. Placebo (placebo q3W for all 8 infusions).
2. Visit windows are ±3 days for Weeks 3, 6, 9, 12, 15, 18 and 21 of double-masked treatment period and Weeks 27, 30, 33, 36, 39, 42 and 45 of open-label treatment period, and ±7 days for Week 24 of double-masked treatment period, Week 48 of open-label treatment period and 30-day Follow-up Period.
3. All patients will be contacted by phone/email the day after infusion for the first and second infusions during the double-masked treatment period and during the open-label treatment period, and thereafter as deemed appropriate; additional phone/email contacts will occur the day after any clinic visit where a patient experiences an infusion-related adverse event.
4. Patients who are proptosis non-responders at Week 24 of the double-masked treatment period will be offered the option to receive 8 infusions of TEPEZZA (10 mg/kg on Day 1 followed by 20 mg/kg for the remaining 7 infusions) in an open-label fashion.
5. All patients will be contacted via phone or email 30 days (±7 days) following the Week 24 (responders and non-responders who choose not to receive open-label TEPEZZA) or Week 48 (non-responders who receive open-label TEPEZZA) Visit.

3.2 Study Endpoints

3.2.1 Primary Endpoint

The primary efficacy endpoint is the change from Baseline at Week 24 in proptosis in the study eye.

3.2.2 Other Efficacy Endpoints

1. The proptosis responder rate (percentage of patients with a ≥2-mm reduction from Baseline in proptosis in the study eye, without deterioration [≥2-mm increase] of proptosis in the fellow eye) at Week 24.
2. The change from Baseline at Week 24 in the GO-QoL questionnaire appearance and visual functioning subscales.

CCI [redacted]

3.2.3 Pharmacokinetic and Anti-drug Antibody Endpoints

1. The peak and trough concentrations of TEPEZZA.
2. The anti-drug antibody (ADA) incidence and titers.

3.2.4 Safety and Tolerability Endpoints

1. The incidence of treatment-emergent adverse events (TEAEs,) serious adverse events (SAEs), TEAEs resulting in premature discontinuation of treatment and AESIs (infusion reactions, hyperglycemia, hearing impairment, new onset inflammatory bowel disease and exacerbation of inflammatory bowel disease).
2. The incidence of ≥Grade 3 TEAEs.
3. The change from Baseline to each scheduled visit in vital signs (blood pressure, heart rate, respiratory rate and temperature).
4. The results of best-corrected visual acuity.
5. The incidence of ≥Grade 3 fasting glucose values.
6. The change from Baseline to each scheduled visit in laboratory evaluations (hematology and chemistry [including thyroid panel and HbA1c]).

3.2.5 Exploratory Endpoints

3.3 Study Eye

At the Baseline (Day 1) Visit, the "study eye" (i.e., the eye with most significant proptosis) will be identified. If both eyes are affected equally, the Investigator will choose the "study eye." Both eyes will be assessed for efficacy, but the study eye will be used to assess the primary efficacy endpoint.

4 Type of Planned Analysis

4.1 Interim Analysis

No interim analyses are planned.

4.2 Primary Analysis

The primary analysis of efficacy, safety and tolerability is planned after all patients complete the Week 24 visit assessments in the double-masked treatment period or terminate early from the study prior to Week 24. Efficacy analyses are based on data collected from the Screening visit through Week 24 visit. Safety analyses are based on data collected up to last patient's Week 24 visit date.

4.3 Follow-up Analysis

The follow-up analysis is planned at the end of the trial after all patients complete the 30-day follow-up period or terminate early from the study during the follow-up period. The follow-up analysis will focus on the open-label treatment period. For the open-label treatment period,

1. Only descriptive summaries will be provided and no statistical modeling or statistical hypothesis tests will be performed.
2. Neither sensitivity nor supplementary analysis will be performed.
3. Analysis results will be presented by the treatment patients received during the double-masked treatment period.

Efficacy analyses are based on data collected from open-label treatment period. Safety analyses are based on data collected from open-label treatment period as well as 30-day follow-up data that are not included in the primary analysis.

5 General Statistical Considerations

The primary efficacy endpoint will be tested at the level of significance of 0.05 two-sided. Other efficacy endpoints will be tested in a sequential order as listed in Section 3.2.2 only when statistical significance is achieved for the primary efficacy endpoint and the other preceding endpoint in the prespecified order is statistically significant at the 0.05 significant level (two-sided).

In general, continuous data will be summarized by treatment group using descriptive statistics (number, mean, standard deviation (SD), minimum, 25^{th} (Q1) percentile, median, 75^{th} (Q3) percentile, and maximum). Categorical data will be summarized by treatment group using frequency tables (frequencies and percentages). SD will be presented to two decimal places beyond the precision with which the data was captured. Mean, median, Q1, and Q3 will be presented to one decimal places beyond the precision with which the data was captured. Minimum and maximum will be presented to the precision with which the data was captured. When count data are presented, the percentage will be suppressed when the count is zero in order to draw attention to the non-zero counts. A row denoted “Missing” will be included in count tabulations where specified on the shells to account for dropouts and missing values. The denominator for all percentages will be the number of patients (including the missing category) in that treatment group within the analysis set of interest, unless otherwise specified. P-values will be rounded to 4 decimal places. If a p-value is less than 0.0001 it will be reported as “<0.0001.” If a p-value is greater than 0.9999 it will be reported as “>0.9999”.

All study-related raw data that support the corresponding tables and figures will be presented in data listings. Unless otherwise noted, all data listings will be sorted by treatment group and patient number.

5.1 Sample Size

A total of 57 patients (38 in the TEPEZZA group and 19 in the placebo group) will be enrolled in the trial to detect at least a 2-mm mean difference between the 2 treatment groups in the change from Baseline of proptosis values at Week 24 in order to have 81% power at the two-sided 0.05 level of significance. The sample size was determined assuming that the mean difference in

proptosis change between the 2 groups is at least 2.0 mm (clinically relevant) and the standard deviation of proptosis change values is 2.5 for both groups (larger than the observed in Phase 2 and Phase 3 active TED trials).

5.2 Randomization, Stratification, and Masking

Patients will be randomly assigned to receive TEPEZZA or placebo using a 2:1 allocation ratio on Day 1 of the double-masked treatment period. No stratification factors will be considered for patient randomization.

In the double-masked treatment period, the clinical trial will be performed in a double-masked manner, while the pharmacists or designees responsible for preparing the TEPEZZA or placebo solutions for IV administration will not be masked to the identity of the trial drug. Pharmacists/designees will provide trial drug in infusion bags (fully diluted for administration) to site personnel with appropriate masked labels. The patient, Investigator, and all other site personnel will be masked to the treatment being administered.

In the double-masked treatment period, all investigative site staff directly involved in this trial, except for unmasked pharmacy personnel, will remain masked from Screening through analysis of the open-label treatment data and all site close-out visits. The Sponsor and its designees may be unmasked after the double-masked treatment period database lock.

5.3 Analysis Set

5.3.1 Intent-to-Treat (ITT)

The ITT analysis set is comprised of all patients who are randomized to receive study drug (either TEPEZZA or placebo). This will be the primary analysis set for efficacy analysis for each treatment period. Patients will be analyzed according to their randomized treatment assignment.

5.3.2 Modified Intent-to-Treat (mITT)

The modified intent-to-treat (mITT) analysis set will include all ITT patients who receive at least one dose of study drug and have at least one post-baseline measurement of the primary efficacy endpoint. The mITT analysis set will be used to perform a supplementary analysis of the primary efficacy analysis for the double-masked treatment period. Patients will be analyzed according to their randomized treatment assignment. If the ITT analysis set is the same as the mITT analysis set because all randomized patients have met the mITT definition, then analysis results based the mITT analysis set will not be provided.

5.3.3 Per Protocol (PP)

The per protocol (PP) analysis set includes all mITT patients who complete the double-masked treatment period and do not incur any major protocol violations that would challenge the validity of their data. The PP analysis set will also be used to conduct as a supplementary analysis of the primary efficacy analysis. Patients will be analyzed according to treatment received. Major protocol violations that challenge the validity of the data will be identified by a masked review

on a regular basis while the study is ongoing and the review for major protocol deviations will be completed prior to the Week 24 database lock for the primary analysis.

5.3.4 Safety

The safety analysis set will include all patients who receive at least one dose of study drug. This will be the primary analysis set for safety evaluation for each treatment period. Patients will be analyzed according to treatment received. In the event a patient receives more than one treatment, the patient will be summarized by the treatment received most frequently.

5.3.5 PK

The PK analysis set will include all patients who receive at least one dose of TEPEZZA and have at least one post-dose PK sample. All PK analyses will be performed on the PK analysis set for each treatment period.

5.4 Assessment Windows

5.4.1 Study Day and TEPEZZA Study Day

The following two definitions of study day will be used for analysis purposes.

- Study Day will be calculated relative to the first dose of study drug (TEPEZZA or Placebo) (Day 1). Study day will be used for all analysis involving study day calculation in either double-masked treatment period or open-label treatment period.
- TEPEZZA Study Day will be calculated relative to the first dose of TEPEZZA whether it occurs in the double-masked treatment period or open-label treatment period (TEPEZZA Day 1). TEPEZZA study day will be used to calculate study day in the open-label treatment period only.

When study day or TEPEZZA study day is used for display or in comparisons the following algorithm will be used:

- Study day (or TEPEZZA study day) = date of assessment - Day 1 (or TEPEZZA Day 1) +1, if date of assessment ≥ Day 1 (or TEPEZZA Day 1).
- Study day (or TEPEZZA study day) = date of assessment - Day 1 (or TEPEZZA Day 1), if date of assessment < Day 1 (or TEPEZZA Day 1).

Note that the day before Day 1 is Day -1 (for analysis, there is no Day 0 for study day).

5.4.2 Baseline and TEPEZZA Baseline

The following two definitions of Baseline will be used for analysis purposes:

- Baseline will be defined as the last observation before the first dose of study drug (TEPEZZA or Placebo). Baseline will be used for all applicable analyses during the double-masked treatment period.

- TEPEZZA Baseline will be defined as the last observation before the first dose of TEPEZZA whether it occurs in the double-masked treatment period or open-label treatment period. TEPEZZA Baseline will be used for applicable analyses during the open-label treatment period.

Change from Baseline (or TEPEZZA Baseline) is defined as the post-baseline value minus the Baseline (or TEPEZZA Baseline) value for the given assessment.

5.4.3 Visit Windows for Analysis

For analyses that present visit-based data, the variables will be summarized based on the scheduled visits with derived analysis visit windows, with the exception of PK data, anti-drug antibody (ADA)/neutralizing antibody (NAb) data, and CCI [redacted], which will be analyzed based on nominal visits. For the visits on and after the first dose date, the actual visit date will be mapped to the derived analysis visit windows based on the study day.

Visit windows have been constructed so that every observation (unscheduled visits included) collected can be allocated to the scheduled analysis visit for planned analyses. The actual assessment day will be mapped to the windows defined for each scheduled study visit with following rules.

For multiple non-missing continuous measurements:

- If more than 1 assessment falls within a visit window, the closest non-missing valid assessment to the scheduled day will be used in the analysis.
- If 2 non-missing assessment actual dates are equidistant from the target day, the later visit will be used in the analysis.
- If there is more than 1 record on the selected day and the time is not recorded, the average will be taken, unless otherwise specified.
- For retest values of laboratory data, the retest value (the last valid observation assessed on the same visit day) will be chosen.

For multiple non-missing categorical measurements:

- For Baseline, the last available record on or prior to the date of the first dose of study drug will be selected. If there are multiple records with the same time or no time recorded on the same day, the value with the lowest severity will be selected (e.g., normal will be selected over abnormal for hematology).
- For postbaseline visits, if there are multiple records with the same time or no time recorded on the same day, the value with the worst severity within the window will be selected (e.g., abnormal will be selected over normal for hematology).

The Analysis Visit Window for the primary endpoint (change of proptosis), other endpoints (CCI [redacted]) and exploratory endpoint (CCI [redacted]) is shown in **Table 1**. The details for the other analysis visit window rules will be included in the TLF shell document.

Table 1 Analysis Visit Windows for Primary Endpoint

Analysis Visit	Target Study Day of Visit	Analysis Visit Window
Baseline	1	Non-missing assessment on or before the first dosing date (randomization)
Week 3	22	Day 2 to 32
Week 6	43	Day 33 to 64
Week 12	85	Day 65 to 106
Week 18	127	Day 107 to 148
Week 24	169	Day 149 to 179*
Week 27	190	Day 180 to 200
Week 30	211	Day 201 to 232
Week 36	253	Day 233 to 274
Week 42	295	Day 275 to 316
Week 48	337	≥ Day 317

* For patients who did not enter the open-label treatment period, Week 24 analysis visit window is ≥ Day 149.

Data from all assessments (scheduled and unscheduled), including multiple assessments, will be included in listings.

5.5 Missing Data Algorithms

In general, missing data will not be imputed unless methods for handling missing data are specified.

5.5.1 Birth Dates

If the date of birth is only partially available, the first of the month will be imputed for any missing days and January will be imputed for any missing months. Imputed dates will not be presented in the data listings.

5.5.2 Medical History Diagnosis Dates

If the onset date of Graves' disease or TED is only partially available, the first of the month will be imputed for any missing days and January will be imputed for any missing months. Imputed dates will not be presented in the data listings.

5.5.3 Adverse Events

For adverse events with incomplete dates, the following rules will be used to impute start and/or stop dates for the sole purpose of determining whether if an AE is treatment emergent. Imputed dates will not appear in the data listings.

For partial or missing AE start dates:

- If day is missing, and the month and year match the month and the year of the first dose date, the day of the first dose date will be imputed and the AE will be considered treatment emergent. Otherwise, the first of the month will be used and the treatment-emergent status will be assessed relative to the dosing start date.
- If the day and month are missing, and the year matches the year of the first dose date, the month and the day of the first dose date will be imputed, and the AE will be considered treatment emergent. Otherwise, January 1 will be used and the treatment emergent status will be assessed relative to the dosing start date.
- If the start date is completely missing, the AE will be considered treatment emergent unless the stop date is complete and prior to the first dose or provides enough partial information to rule out a treatment-emergent status.
- If the stop date is complete and the imputed start date is after the actual stop date, then the start date will be imputed as the stop date and the treatment-emergent status will be assessed relative to the dosing start date.

Any missing intensity assessments for AEs will be imputed as "severe" and any missing relationship to study drug will be considered "related" for summary purposes.

5.5.4 Medication Dates

For prior and concomitant medications with incomplete dates, the following rules will be used to impute start and/or stop dates for the purposes of determining if a medication is prior or concomitant only. Imputed dates will not be presented in the data listings.

For partial or missing start dates:

- If day is missing, and the month and year match the month and the year of the first dose date, the day of the first dose date will be imputed. Otherwise, the first of the month will be used.
- If day and month are missing and the year matches the year of the first dose date, the month and the day of the first dose date will be imputed. Otherwise, January 1 will be used.
- If the start date is completely missing, the start date will not be imputed. If the stop date is after first dose date, the medication will be considered to be concomitant. If the stop date is prior to the first dose date, the medication will be considered to be prior.
- If the stop date is complete and the imputed start date is after the actual stop date, then the start date will be imputed as the stop date.

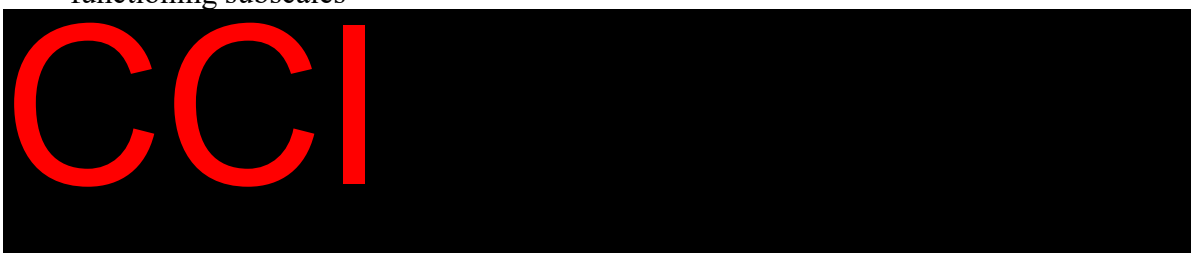
For partial or missing stop dates:

- For stop dates, if the day is missing, then the last day of the month will be used.
- If the month is missing, then December will be used.
- If the stop date is completely missing, then the latest date available in the dataset for the patient will be used.
- All medications with a missing stop date will be considered as concomitant.

5.6 Multiplicity

If a statistically significant result in favor of TEPEZZA is achieved on the primary efficacy endpoint, the other efficacy endpoints (with the primary analyses defined in the study) will be analyzed in a hierarchical manner to control for multiplicity. For each other efficacy endpoint, TEPEZZA will be tested against placebo at the two-sided 0.05 significance level only if the test statistic was statistically significant in favor of TEPEZZA for the endpoint preceding it in the following hierarchical order:

1. Proptosis responder rate (the percentage of patients with a ≥2-mm reduction from Baseline in proptosis in the study eye, without deterioration [≥2-mm increase] of proptosis in the fellow eye) at Week 24
2. Change from Baseline at Week 24 in the GO-QoL questionnaire appearance and visual functioning subscales



6. Patient Disposition

6.1 Disposition

The counts of patients who were screened will be presented. The counts and percentages of patients who are screen failures, randomized, and included in each analysis set will be presented.

Patients who completed or discontinued from the treatment period and study will be presented based on the number of patients who were randomized in each treatment group and entered in each treatment period along with both treatment groups pooled. Reasons for discontinuation for each treatment period or from the study will be summarized separately by treatment group for each treatment period. In addition, whether patients discontinued from the treatment period or study due to COVID-19 will also be summarized separately by treatment group for each treatment period.

6.2 Protocol Deviations

Patients with any major protocol deviations will be tabulated for each treatment group by deviation category (Consent/Assent; Procedures/Documentation; Eligibility; Study Medication Dispensing/Dosing; Other Medication Dosing; Prohibited Action; Serious Adverse Event

Reporting; Visit /Assessment Schedule) and treatment period for ITT analysis set. All major and minor protocol deviations observed in the trial will be listed by-patient for all patients enrolled.

7. Demographics and Baseline Characteristics

Demographics and baseline characteristics collected at Screening will be summarized for the ITT analysis set, mITT analysis set, and safety analysis set for the double-masked treatment period, and ITT analysis set for the open-label treatment period. If mITT and safety analysis sets are the same as the ITT analysis set, then summaries based on mITT and safety analysis sets will not be presented. Continuous variables (i.e., age [in years], baseline weight [kg], baseline height [cm], baseline body mass index [BMI in kg/m^2], and time since diagnosis of TED [in years]) will be summarized by treatment group and overall using descriptive statistics. For the similar summary of the open-label treatment period, TEPEZZA Baseline will be used to summarize the baseline weight, baseline height and baseline BMI, and time since diagnosis of TED. The following categorical variables will be summarized by reporting the number and percentage of patients in each category for each treatment group and overall.

- Age category (<65, ≥65 years old)
- Sex (Male, Female)
- Study eye (right, left)
- Childbearing potential (Yes, No or N/A)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, and Other)
- Tobacco Use (Former, Current, Never)
- Alcohol Use (Former, Current, Never)

Individual patient's demographics and baseline characteristics will be presented in a by-subject listing for all patients enrolled in the trial.

Ocular baseline characteristics will be summarized for the ITT analysis set, mITT analysis set, and safety analysis set during the double-masked treatment period and for the ITT analysis set during the open-label treatment period, on the study eye and fellow eye, separately. If mITT and safety analysis sets are the same as the ITT analysis set, then summaries based on mITT and safety analysis sets will not be presented. For the open-label treatment period, baseline values will be recalculated based on the TEPEZZA Baseline. Continuous variables for patient ocular baseline characteristics are proptosis (measured as proptosis evaluation of the Clinical Measures of Severity using a Hertel instrument provided by the Sponsor for consistency in measurement), CCI and clinical activity score (CAS). These continuous

variables will be summarized using descriptive statistics for each treatment group and overall. Categorical variables such as binocular diplopia (measured as part of the Clinical Measures of Severity) and best corrected visual acuity (BCVA) (Normal/Mild/Moderate Low Vision/Severe Low Vision/ Blindness) will be summarized by treatment group and overall for each category.

7.1 Substance Use

The data of substance use will be presented in a by-subject listing for all patients enrolled in the trial.

7.2 Medical History

General medical history will be summarized by system and organ class (SOC) and preferred term (PT) coded according to Medical Dictionary for Regulatory Activities (MedDRA) version 24.0.

The medical history data for all patients enrolled in the trial will be presented in a by-subject listing.

7.3 Inclusion and Exclusion Criteria

Inclusion and exclusion criteria deviations noted in the electronic case report form (eCRF) will be presented for all patients who provide informed consent in a by-subject listing. If there are patients who were enrolled in the trial despite they did not meet at least 1 of the entry eligibility criteria, the eligibility criterion or criteria that patients did not meet will be summarized by treatment group.

8. Treatments and Medications

8.1 Prior and Concomitant Medications

The concomitant medications will be summarized based on the eCRF collected information for all medications that patients have taken before or during the course of the clinical trial.

Prior medications are defined as all medications which start before the first dose of study drug. Concomitant medications during the double-masked treatment period are defined as all medications that start on or after the first dose of study drug until 21 days following the last dose of study drug or medications that start before the first dose of study drug and continue during the double-masked treatment period. Medications that start before the first dose of study drug and continue during the double-masked treatment period will be counted as both prior and concomitant medications. Concomitant medications during the open-label treatment period are defined as all medications that start on or after the first dose of TEPEZZA until 21 days following the last dose of TEPEZZA in the open-label treatment period, or medications that start before the first dose of TEPEZZA in the open-label treatment period and continue during the open-label treatment period. Medications that start in the double-masked treatment period and continue in the open-label treatment period will be counted as concomitant medications for both double-masked treatment period and open-label treatment period. Concomitant medications

during the follow-up period are defined as all medications that start after 21 days of the last dose of study drug through completion of follow-up period, or medications that start before 21 days of the last dose of study drug and continue during the follow-up period.

All medications will be mapped to the preferred terms according to WHODrug dictionary version March 2021.

The number and percentage of patients in each treatment group using prior medications and concomitant medications will be summarized by Anatomical Therapeutic Chemical (ATC) Level 4 term and Preferred Term. At each level of report, a patient will be counted only once even if the patient has taken more than one medication in the level.

All prior and concomitant medications (including concomitant medications during the follow-up period) will be presented in a by-subject listing with an indicator to identify whether their use is prior and /or concomitant.

8.2 Concomitant Procedures

All concomitant procedures performed during the trial will be presented in a by-subject listing for all patients enrolled.

8.3 Restricted Medications and Surgical Procedures

The restricted medications and surgical procedures are specified in Table 9.1 in protocol.

- Restricted surgical procedures are (1 category):
  - "orbital irradiation, orbital decompression or strabismus surgery". This category will be captured via medical history eCRF form.
- Restricted medications or drugs are (9 categories):
  - "steroids", "Rituximab (Rituxan® or MabThera®)", "Tocilizumab (Actemra® or Roactemra®)", "TEPEZZA", "monoclonal antibody", "non-steroid immunosuppressive agent (other than rituximab or tocilizumab)", "investigational agent" and "Neonatal Fc receptor (FcRn) agents". These categories will be captured via concomitant medications eCRF form.
  - "illicit drug/alcohol abuse": This category will be captured via medical history and adverse event eCRF forms.

All restricted medication categories described above are reviewed manually by Horizon patient safety and pharmacovigilance.

The number and percentage of patients who received restricted surgical procedures in each treatment group will be summarized by SOC and preferred terms.

The number and percentage of patients who received restricted medications in each treatment group for each treatment period will be summarized by

- ATC Level 4 term and preferred terms for concomitant medications.

- SOC and preferred terms for medical history and adverse event

A patient who reported the same event more than once will be counted only once in the applicable category, SOC (or ATC Level 4), and preferred term for that event.

All restricted medications and surgical procedures will be presented in a by-subject listing.

8.4 Study Treatments

8.4.1 Extent of Exposure

Extent of exposure will be measured and summarized by duration of exposure and total number of infusions and presented by treatment period (double-masked treatment period/open-label treatment period) and TEPEZZA.

The duration of exposure (days) for each treatment period will be calculated as:

The duration of exposure (days) for each treatment period = the last dosing date in each treatment period - the first dosing date in each treatment period +1.

The duration of exposure (days) to TEPEZZA will be calculated as the sum of treatment duration to TEPEZZA in double-masked treatment period and open-label treatment period.

The total number of infusions given is the sum of actual infusions received by patient for each treatment period.

The total number of infusions to TEPEZZA is the sum of TEPEZZA infusions received in double-masked treatment period and open-label treatment period

8.4.2 Treatment Compliance and Modifications

All doses of study drug will be administered and recorded by study site personnel, and any deviations from the dosing schedule will be entered into the patient's eCRF. The study drug administration records will be listed for each patient with study drug administered, start date and time, stop date and time, dose, if there's any dose interruptions incompleteness and reasons.

Summaries will be provided for the count and percentages of patients with each of the following by treatment group for each treatment period:

- Planned doses that were not administered completely
- Infusion interruptions

Planned TEPEZZA doses that were not administered completely and infusion interruptions to TEPEZZA in double-masked treatment period and open-label treatment period will be summarized by treatment group.

9. Efficacy Analysis

Efficacy analysis will be conducted for the double-masked period (as the primary analysis) and open-label period (as the follow-up analysis) separately based on the ITT analysis set. Efficacy analysis results from the primary analysis (i.e., based on the data collected during the double-masked period) will be used for comparison between the two treatment groups. Both study and fellow eyes will be used for the efficacy analyses during the primary analysis, but the statistical inference will be made based on study eye analysis results only. For the follow-up analysis, which will be conducted using subsequent assessments collected during the open-label treatment period, only descriptive summaries of efficacy analysis results will be provided.

9.1 Intercurrent Event Handling Strategy for Estimands

Potential intercurrent events (ICEs) like discontinuing from the treatment may occur. Patients who discontinue from the treatment will be encouraged to stay in the trial after those ICEs and data will be collected through the planned study completion visit.

In general, treatment policy strategy will be used to account for the ICEs for defining estimands corresponding to continuous efficacy endpoints, including the primary efficacy endpoint. All efficacy analyses will be based on the observed data except for categorical efficacy endpoints. For categorical endpoints, composite strategy will be used. For binary responder endpoints, patients whose Week 24 assessment is missing will be considered treatment failures (non- responders). For ordinal response endpoint, missing value at Week 24 will be imputed using the last assessment value available for patients whose Week 24 assessment is missing.

9.2 Primary Efficacy Endpoint

The primary efficacy endpoint is the change in proptosis from Baseline at Week 24 in the study eye.

Population	ITT Analysis Set defined as all patients who were randomized in the trial
Endpoint	Change from Baseline at Week 24 in proptosis of the study eye
ICE	Discontinued from treatment early
Population-level Summary	Difference (TEPEZZA minus placebo) in least squares means of changes from Baseline at Week 24 in proptosis of the study eye between the treatment groups

9.2.1 Primary Analysis Method

The primary efficacy endpoint analysis will be conducted based on the ITT analysis set.

A Mixed-Model for Repeated-Measures (MMRM) analysis of covariance model fitting to the individual change from Baseline values for the study eye will be used for the analysis of change from Baseline in proptosis. The model includes Baseline value, treatment group, visit, visit-by-

treatment and visit-by-Baseline value as fixed effects, and patient as a random effect as shown below:

$$y_t = X\beta + Z\gamma + \epsilon,$$

where

- y_t is a vector of change in proptosis from Baseline at visit t
- visit t is Week 3, 6, 12, 18 or 24
- β is an unknown vector of fixed-effects parameters with known design matrix X = (Baseline value, treatment group, t, t *treatment group, t * Baseline value)
- γ is an unknown vector of random-effects parameters with known design matrix Z = patient, and
- ϵ is an unknown random error vector.

It is assumed that γ and ϵ are Gaussian random variables that are uncorrelated and have expectations 0 and variances G and R, respectively. The variance of y is

$$V = ZGZ' + R,$$

where G and R are unstructured covariance.

If there are model convergence issues with an unstructured variance-covariance matrix, the following three variance-covariance matrices will be attempted in order until one converges: heterogeneous Toeplitz (TOEPH), heterogeneous compound symmetry (CSH), and Toeplitz (TOEP). If there are any patients in the ITT analysis set without post-baseline values, a change from Baseline value of 0 will be imputed at the first post-baseline visit (in order to avoid exclusion of these patients from the MMRM analysis).

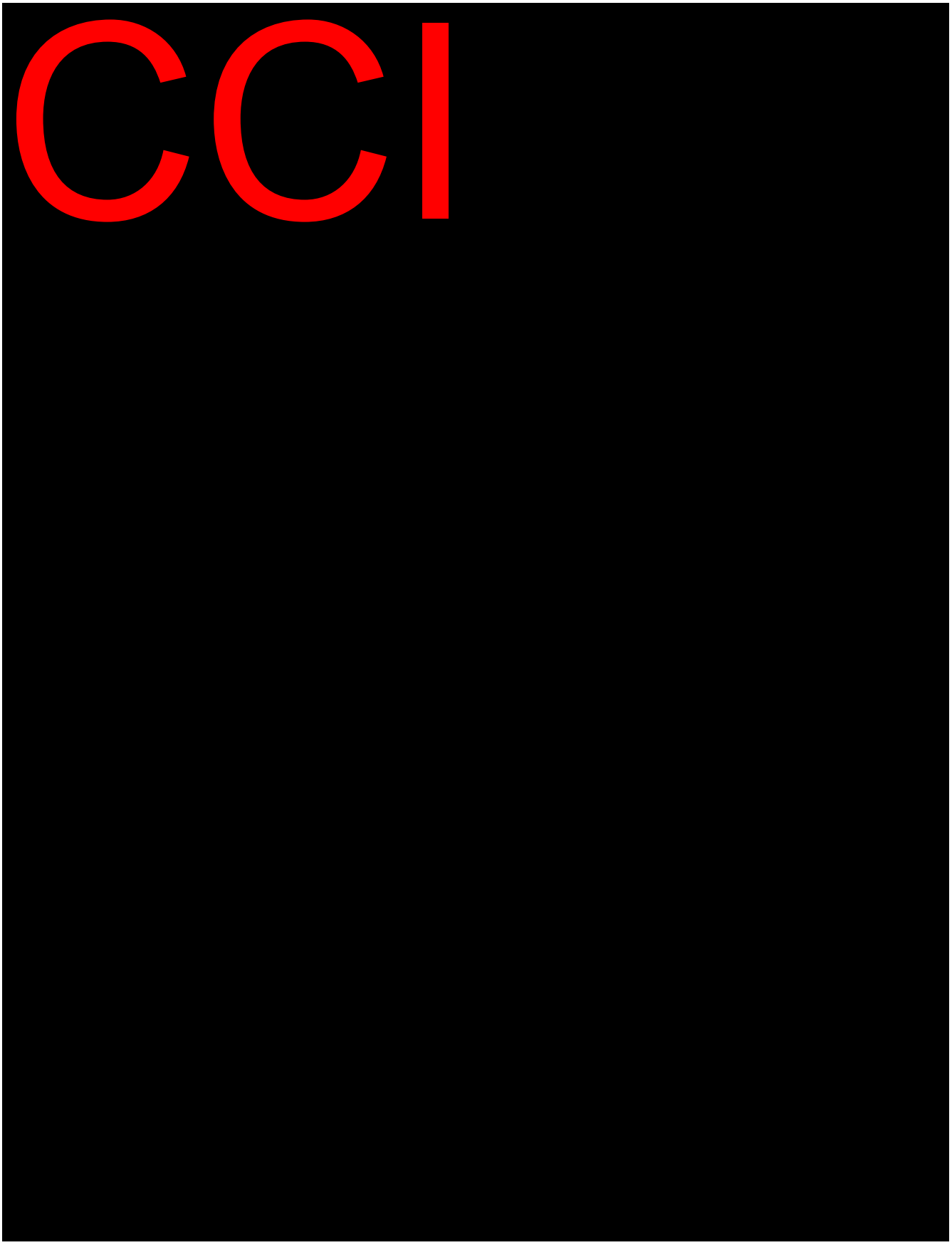
The p-values for all terms in the model will be presented, as well as the overall treatment group least squares (LS) means and associated standard errors (SE), and their difference, SE of the difference, 95% confidence intervals (CIs) for the difference and p-value. The primary analysis will be to test the treatment difference at Week 24. Analysis results at Week 24 present the estimated LS means, SEs and their difference with the SE, 95% CI for the difference and p-value.

Further, the analysis will be repeated using assessments obtained from the fellow eye based on the ITT analysis set.

Descriptive summaries of change from Baseline in the proptosis for study eye and fellow eye at each scheduled visit will be provided by treatment group for each treatment period based on the ITT analysis set.

Changes from Baseline in proptosis of the study eye over time during the double-masked treatment period will be depicted by treatment group using LS means of change and corresponding 95% CI.

9.2.2 Sensitivity Analysis Method


CCI

9.2.3 Supplementary Analysis

The primary efficacy analysis described in Section 9.2.1 will be repeated using the mITT and PP analysis sets on study eye only.

Descriptive summaries of change from Baseline in the proptosis for study eye and fellow eye at each scheduled visit will be provided by treatment group for each treatment period using the mITT and PP analysis sets.

9.2.4 Subgroup Analysis

The primary efficacy endpoint will be evaluated for subgroups of interest, including:

- Race (White, Black or African American, Asian, Other)
- Tobacco use status (Non-user, User)
- Age (<65 years, >=65 years)
- Sex (Male, Female)

For race, if the number of patients in any of the race groups is less than 5, we will pool “Black or African American”, “Asian”, “Other” into “Non-White”. For other variables, if the number of patients is less than 5 in any subgroup, then subgroup analysis will not be performed. For tobacco use status, tobacco uses of “Never” or “Former” will be grouped as tobacco “Non-users”, and tobacco uses of “Current” will be considered tobacco “Users” on the substance use eCRF.

9.2.5 Exploratory Analysis

9.3 Other Efficacy Endpoints

Inferential testing on other efficacy endpoints will be made only when the primary efficacy endpoint reaches statistical significance at 0.05 (two-sided) in favor of TEPEZZA. Other efficacy endpoints (with the main analyses defined in the study eye as appropriate) will be analyzed using the ITT analysis set in a hierarchical manner presented in the order in Section 5.6. For each outcome measure, TEPEZZA will be tested against placebo at the two-sided 0.05 significance level only if the test statistic was statistically significantly in favor of TEPEZZA for the outcome measure preceding it in the hierarchical order.

9.3.1 Proptosis Responder Rate at Week 24

The proptosis responder rate is defined as the percentage of patients with a ≥2-mm reduction from Baseline in proptosis in the study eye, without deterioration [≥2-mm increase] of proptosis in the fellow eye at each scheduled visit.

Population	ITT Analysis Set defined as all patients who were randomized in the trial
Endpoint	Proportion of patients who have achieved the proptosis response rate at Week 24. The proptosis responder rate at Week 24 is defined as the percentage of patients with a ≥2-mm reduction from Baseline in proptosis in the study eye, without deterioration (≥2-mm increase) of proptosis in the fellow eye at Week 24.
ICE	Discontinued from treatment early
Population-level Summary	Difference (TEPEZZA minus placebo) in proptosis responder rates at Week 24 between the treatment groups

9.3.1.1 Primary Analysis Method

Patients whose Week 24 evaluation is missing will be considered treatment failures (non-responders). Difference in proptosis responder rates at Week 24 for study eye between the two treatment groups (TEPEZZA minus placebo) on the ITT analysis set will be compared using a two-sided Fisher Exact test. A 95% exact CI around the proportion difference between the two treatment groups will be provided. The same analysis will be repeated using observed results. In this analysis, only patients with a non-missing evaluation at Week 24 are included.

Further, the definition of responder will be applied for each scheduled visit in double-masked treatment period and the analysis of risk difference will be performed, considering patients missing the evaluations as treatment failures (non-responders) and separately considering the observed results only. In order to investigate the effect of treatment on the fellow eye, the analysis for fellow eye will be conducted. The same analyses will be repeated to include

responder status at each scheduled visit in open-label treatment period, but no analysis of risk difference will be performed.

Line plot of proptosis responder rate over time by treatment group for the study eye based on the ITT analysis set will be provided for each treatment period.

9.3.1.2 Sensitivity Analysis Method

CCI

-
-
-
-

9.3.2 Change from Baseline at Week 24 in the GO-QoL Questionnaire Appearance and Visual Functioning Subscales

The GO-QoL questionnaire (see Protocol section 17.5) has CCI

9.3.3 CCI

CCI

-
-
-
-


CCI

The same primary analyses described in Section 9.3.1.1 will be repeated for this endpoint.

9.4 Exploratory Endpoints

9.4.1 CCI [redacted]

CCI [redacted]

10. Safety Analysis

Safety parameters include monitoring of TEAEs, SAEs, vital sign measurements, laboratory tests and visual acuity.

Safety data will be summarized and listed with no statistical hypothesis testing performed. Continuous variables will be summarized using the mean, the SD, median, 25th percentile, 75th percentile, minimum value, and maximum value. Categorical variables will be summarized using

frequency counts and percentages. All safety summaries and analyses will be conducted separately for each treatment period based on the safety analysis set.

10.1 Adverse Events

Adverse events will be coded using MedDRA version 24.0. AEs that occur during Screening and prior to dosing on Day 1 will be considered pre-treatment AEs. TEAEs for the double-masked treatment period is defined as AEs with onset date on or after the first dose of study drug during the double-masked treatment period prior to receiving the first dose of TEPEZZA during open-label treatment period for patients who entered open-label treatment period, or through 3 weeks (21 days) after the last dose of study drug during the double-masked treatment period for patients who did not enter open-label treatment period. TEAEs for open-label treatment period defined as AEs with onset date on or after first dose of TEPEZZA during the open-label treatment period through 3 weeks (21 days) after last dose of TEPEZZA during the open-label treatment period. All analyses for TEAEs will be split to TEAEs reported during the double-masked treatment period and open-label treatment period separately. The follow-up AE reporting period begins 3 weeks (21 days) after the last dose of study drug through completion of the follow-up period.

An overall summary of all TEAEs including the number and percentage of patients experiencing at least one occurrence of a TEAE, serious TEAEs, treatment-related TEAEs, treatment-related serious TEAEs, TEAEs resulting in study drug discontinuation, TEAEs of special interest, TEAEs leading to death along with the summary of National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) grade 3 or higher will be provided.

Additionally, AEs and SAEs reported during the follow-up period after each treatment period will be presented by SOC and PT separately.

All AEs, AEs with CTCAE Grade 3 or higher, SAEs, AEs leading to discontinuation of study drug, AE of special interest, and deaths will be presented in separate by-subject listings.

Signs and symptoms associated with Infusion reaction and anaphylaxis will be recorded and included in a by-subject listing.

10.1.1 Incidence of Adverse Events

The incidence of AEs will be reported by SOC and PT level for each treatment period. The patient will be counted only once in corresponding category if the patient experiences more than one event under the same SOC or PT level.
The total number of TEAEs and the number and percentage of patients with at least one TEAE for each SOC and PT will be presented by treatment group.

Incidence of TEAEs by SOC and PT by the following subgroups for each treatment period will be provided.

- Race (White, Black or African American, Asian, Other)
- Tobacco use status (Non-user, User)
- Age (<65 years, >=65 years)
- Sex (Male, Female)

For race, if the number of patients in any of the race groups is less than 5, we will pool "Black or African American", "Asian", "Other" into "Non-White". For other variables, if the number of patients is less than 5 in any subgroup, then subgroup analysis will not be performed. For tobacco use status, tobacco uses of "Never" or "Former" will be grouped as tobacco "Non-users", and tobacco uses of "Current" will be considered tobacco "Users" on the substance use eCRF.

10.1.2 Relationship of Adverse Events to Study Drug

The relationship between study treatment and AE will be captured on the eCRF. The AE records with missing relationship information will be summarized as related AEs.

Summary of all TEAEs and SAEs respectively will be presented by SOC, PT and relationship (Unrelated/Related) to study drug.

10.1.3 Severity of Adverse Event

The severity of AEs will be evaluated by the NCI-CTCAE (v4.03). the grade will be assessed by the investigator as mild (Grade 1), moderate (Grade 2), severe (Grade 3), life-threatening (Grade 4), or fatal (Grade 5).
For AE grade tables, the patient will be counted once with the highest-grade event if he/she experiences more than one event under given SOC or PT level.

A summary of CTCAE Grade 3 or higher TEAEs by SOC and PT will be presented by treatment group.

Summary of all TEAEs and SAEs respectively will be presented SOC, PT and maximum severity of AEs.

10.1.4 Serious Adverse Events

The SAE is defined as any event that results in death, is immediately life threatening, requires inpatient hospitalization or prolongation of existing hospitalization, results in persistent or significant disability/incapacity, or is a congenital anomaly/birth defect.

A summary of all SAEs by SOC and PT will be presented by treatment group.

10.1.5 Adverse Events Leading to Treatment Discontinuation

AEs leading to treatment discontinuation will be captured in the eCRF AE page. Besides treatment discontinuation, other consequences include drug interrupted, dose reduced, delayed, and increased will also be collected.

A summary of TEAEs leading to premature discontinuation of treatment by SOC and PT will be presented by treatment group.

10.1.6 Adverse Events of Special Interest (AESI)

AESIs include infusion reactions, hyperglycemia, hearing impairment, new onset inflammatory bowel disease and exacerbation of inflammatory bowel disease.

Infusion reactions will be collected on the AE eCRF form; hyperglycemia will be captured via standard MedDRA query (SMQ) narrow term "Hyperglycaemia/new onset diabetes mellitus"; hearing impairment will be captured via sub-SMQ term "Hearing impairment" (under SMQ “Hearing and vestibular disorders”) and high level term (HLT) term "Hearing losses"; and new onset inflammatory bowel disease and exacerbation of inflammatory bowel disease will be captured via HLT term “Colitis (excl infective)”.

A summary of TEAEs of special interest by SOC and PT will be presented and corresponding AESI listing will be provided

10.2 Clinical Laboratory Evaluations

With the exception of urine pregnancy tests, a central laboratory will be used for all protocol-specified clinical laboratory parameters. Laboratory results reported in conventional units will be converted to International System of Units (SI) for all summaries and listings for consistent results presentation.

Safety laboratory including hematology, chemistry, thyroid function tests and HbA1c values and change from Baseline values will be summarized by visit and treatment group for each treatment period using descriptive statistics. For fasting glucose and HbA1c, means and 95% CI of the test results measured at each scheduled visit will be plotted.

The laboratory assessment will be categorized as low, normal or high based on normal ranges. Shift tables using categories of low, normal and high from Baseline (i.e., based on normal reference ranges) to each post-baseline visit will be generated by treatment group for each treatment period for hematology, chemistry parameters, thyroid function and HbA1c, respectively.

The laboratory assessment will be graded using the NCI-CTCAE (v4.03) grading scale, when available. A shift table for glucose by NCI-CTCAE grade and visit will be generated by treatment group. Another table with summary of incidence of ≥ Grade 3 glucose will be provided.

For those patients who experienced hyperglycemia (per AE reporting), their safety laboratory assessment results during the double-masked treatment period will be examined and summarized using the similar manner described above.

10.3 Vital Sign and Weight Measurements

Vital signs measurements will include systolic and diastolic blood pressures, heart rate, respiratory rate, height, weight, BMI and temperature.

Descriptive summaries of observed and change from Baseline values will be presented for each vital sign parameter by treatment group and visit for each treatment period.

A summary table for weight loss and a shift table for hypertension by NCI-CTCAE grade will be generated by treatment group and visit for each treatment period.

10.4 Other Safety Data

Pregnancy test will be performed at each scheduled visit and the data will be included in a by-subject listing.

For best-corrected visual acuity (BCVA), shift tables for study and fellow eye each will be presented providing the count of patients with each type of finding (i.e., Normal, Mild, Moderate Low Vision, Severe Low Vision and Blindness) at each assessment visit by treatment group for each treatment period. BCVA findings are categorized as follow:

- Normal: 20/12 to 20/40
- Mild: worse than 20/40 to 20/70
- Moderate Low Vision: worse than 20/70 to 20/200
- Severe Low Vision: worse than 20/200 to 20/400
- Blindness: worse than 20/400

11. Immunogenicity: Anti-drug Antibody

ADA samples will be collected prior to the infusion at scheduled visits. If the ADA test is positive after confirmatory and reactive titer testing, the sample will then be tested for neutralizing antibodies (NAb).

Overall positive ADA result for a patient will be defined as at least one post Day 1 confirmed positive ADA result. Cumulative negative result will be defined as negative ADA result at all available time points for a patient.

The reported titer (ratio) value is the minimum required dilution (1:2) * the dilution factor used.

Titer (Ratio) – reported value	Titer (Result) – converted value
1:<2	2
1:x	X

For example, x could be 4, 10, 40, 64, and 100.

The incidence and titer of positive ADA results will be summarized by visit and treatment group using descriptive statistics for the safety analysis set in each treatment period.

All ADA and NAb test results will be presented in a by-subject listing for the safety analysis set in each treatment period.

12. Pharmacokinetics

Pharmacokinetic samples will be collected pre-dose and at the end of infusion on scheduled visits, Day 1, Weeks 3 and 12 of the double-masked treatment period (all patients) and at Weeks 24, 27 and 36 (non-responders who elect to receive open-label TEPEZZA); a single sample will also be collected at Week 24 (or premature withdrawal [PW1]) of the double-masked treatment period (proptosis responders and proptosis non-responders who elect not to receive open-label TEPEZZA) and Week 48 of the open-label treatment period (non-responders who elect to receive open-label TEPEZZA).

A patient listing of serum concentration-time data will be presented by visit. Teprotumumab peak and trough (i.e., prior to dose) concentrations will be summarized by visit and ADA status (overall positive vs cumulative negative; only to be performed if more than 2 patients receiving TEPEZZA treatment with a positive ADA observed in the trial) using descriptive statistics. The analyses will be based on the PK analysis set in each treatment period.

Further, the mean and CV% (calculated as 100%*SD/mean) of pre-dose serum concentrations will be provided by visit in the subset of patients who are confirmed ADA positive, NAb positive (if applicable) and ADA negative for each visit. This analysis will be provided for the PK Population.

13. Changes in the Planned Analysis

- To ensure the control of the overall Type I error rate in analyzing multiple efficacy endpoints, the process to test individual efficacy endpoints has been added in the SAP section 5 and section 9, as an addition to the protocol.
- TEPEZZA Baseline is defined in the SAP for analysis relative to baseline in the open-label treatment period.
- TEPEZZA study day is defined in the SAP for analysis relative to first dose date during the open-label treatment period.

- To better characterized GO-QoL, analyses of patients with minimal clinically important change in the GO-QoL subscales will be provided.
- Summary and shift tables for chemistry (excluding glucose) will be based on all records (fasting and non-fasting records).

14. References

Campbell G, Pennello G, and Yue L. Missing Data in the Regulation of Medical Devices. Journal of Biopharmaceutical Statistics. 2011;21(2):180-95.

Douglas RS, Kahaly GJ, Patel A, et al. Teprotumumab for the treatment of active thyroid eye disease. New England Journal of Medicine. 2020;382(4):341-52.

Heitjan DF and Little RJA. Multiple Imputation for the Fatal Accident Reporting System. Journal of the Royal Statistical Society, Series C. 1991; 40:13–29.

Lumyongsatien M, Keeratidamkerngsakul B, Pornpanich K, et al. Development and psychometric properties of the Thai Graves' ophthalmopathy quality of life (GO-QOL) questionnaire. Journal of Patient-Reported Outcomes. 2019;4(1): 1.

Rubin DB Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons. 1987.

Schenker N and Taylor JMG. Partially parametric techniques for multiple imputation. Computational Statistics & Data Analysis. 1996; 22:425-446.

Smith TJ, Kahaly GJ, Ezra DG, et al. Teprotumumab for Thyroid-Associated Ophthalmopathy. The New England Journal of Medicine. 2017;376(18):1748-61.

Yan X., Lee S., and Li N. Missing Data Handling Methods in Medical Device Clinical Trials, Journal of Biopharmaceutical Statistics. 2009; 19(6):1085-1098

Yuan Y. Multiple Imputation Using SAS Software. Journal of Statistical Software. 2011; 45(6):1-25.

15. Appendices

15.1 Schedule of Assessments

		Double-masked Treatment Period									Open-label Treatment Period (Proptosis Non-responders who Choose to Receive Open-label TEPEZZA)								30 days[1] after W24 or W48/ EOS/ PW
Trial Visit	**Screening[2]**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9/EOT1/ PW1[3]**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17/EOT2 / PW2[3]**	
Week (W)	**-28 days**	**Day 1/ BL**	**W3**	**W6**	**W9**	**W12**	**W15**	**W18**	**W21**	**W24**	**W27**	**W30**	**W33**	**W36**	**W39**	**W42**	**W45**	**W48**	
Visit Window (±days)			**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±7**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±7**	**±7**
Informed consent	X																		
Review inclusion/exclusion criteria	X	X																	
Demographics	X																		
Medical history and prior medications[4]	X[5]	X																	
Weight[6]	X				X					X			X					X	
Randomization[7]		X																	
Efficacy assessments																			
Clinical Measures of Severity – includes proptosis and CCI [redacted]	X	X[8]	X	X		X		X		X[9]	X	X		X		X		X	
CCI [redacted]		X	X	X		X		X		X	X	X		X		X		X	
CCI [redacted]	X									X								X	
Clinical Activity Score[11]	X	X				X				X				X				X	
GO-QoL Questionnaire		X		X		X				X		X		X				X	
Treatment response assessment[9]										X									
Safety assessments																			
Pregnancy test[12]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	

		Double-masked Treatment Period									Open-label Treatment Period (Proptosis Non-responders who Choose to Receive Open-label TEPEZZA)								30 days[1] after W24 or W48/ EOS/ PW
Trial Visit	**Screening[2]**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9/EOT1/ PW1[3]**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17/EOT2 / PW2[3]**	
Week (W)	**-28 days**	**Day 1/ BL**	**W3**	**W6**	**W9**	**W12**	**W15**	**W18**	**W21**	**W24**	**W27**	**W30**	**W33**	**W36**	**W39**	**W42**	**W45**	**W48**	
Visit Window (±days)			**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±7**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±7**	**±7**
Visual acuity	X[13]	X		X		X		X		X		X		X		X		X	
Vital signs[14]	X	X[14]	X[14]	X	X	X	X	X	X	X[14]	X[14]	X	X	X	X	X	X	X	
Clinical laboratory tests																			
Fasting chemistry	X	X		X		X		X		X		X		X		X		X	
Thyroid function (FT_3, FT_4, TSH)[15]	X	X				X				X				X				X	
Hematology	X	X		X		X		X		X		X		X		X		X	
HbA1c	X[16]	X				X				X				X				X	
ADA/NAb samples[17]		X	X			X				X	X			X				X	
AE, SAE assessment	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Prior/concomitant medications	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Pharmacokinetic samples[18]		X	X			X				X	X			X				X	
Biomarker samples[19]		X	X			X				X[20]									
Photographs[21]		X								X								X	
Trial drug infusion		X	X	X	X	X	X	X	X	X[22]	X	X	X	X	X	X	X		
Contact (phone/email) for safety 24 hours post dose[23]		X[23]	X[23]							X[23]	X[23]								

ADA=anti-drug antibody; AE=adverse event; BL=Baseline; EOS=End-of-Study/Trial; EOT=End-of-Treatment; FT_3=free triiodothyronine; FT_4=free thyroxine; GO-QoL=Graves' Ophthalmopathy Quality of Life; HbA1c=glycated hemoglobin; NAb=neutralizing antibody; PW=premature withdrawal; q3W=once every 3 weeks; SAE=serious adverse event; TSH=thyroid-stimulating hormone; W=week.

Footnotes:

1. All patients will be contacted via phone or email 30 days (±7 days) following the Week 24 (responders and non-responders who choose not to receive open-label TEPEZZA) or Week 48 (non-responders who receive open-label TEPEZZA) Visit.
2. Screening procedures can occur over more than 1 day/clinic visit provided consent is obtained first and all assessments are completed within the designated window.
3. If a patient prematurely discontinues trial drug during the double-masked Treatment Period or the open-label Treatment Period, he/she will return for a clinic visit and undergo the W24/PW1 or W48/PW2 assessments, respectively, with the exception of the collection of blood samples for biomarker evaluations.

4. Medical history will include tobacco, alcohol and other substance use history and thyroid disease history and treatment.
5. Thyroid eye disease must be stable, chronic (inactive) (not progressing, non-sight-threatening but with an appreciable impact on daily life), with thyroid eye disease diagnosed ≥3 years but <8 years prior to Screening.
6. The weight obtained at Screening will be used to calculate dose for Day 1 and Weeks 3, 6 and 9. The weight obtained at Week 9 will be used in dose calculations at Weeks 12, 15, 18 and 21. The dose on Weeks 24, 27, 30 and 33 of the open-label Treatment Period will be based on the Week 24 weight; weight measured at Week 33 will be used in dose calculations at Weeks 36, 39, 42 and 45.
7. On Day 1, patients will be randomized in a 2:1 ratio to receive either: a) TEPEZZA (10 mg/kg on Day 1 followed by 20 mg/kg q3W for the remaining 7 infusions) or b) placebo (q3W for all 8 infusions); baseline assessments will be performed prior to dosing.
8. Patients who have a ≥2-mm decrease from Screening in proptosis in the study eye are not eligible for randomization.
9. Response assessment will include whether the patient is a proptosis responder or non-responder; therefore, the proptosis measurement should be one of the first assessments completed during the Week 24 Visit.
10. CCI [redacted]
11. Clinical Activity Score must be ≤1 in both eyes at the Screening and Baseline Visits.
12. For women of childbearing potential(including those with an onset of menopause <2 years prior to Screening, non-therapy-induced amenorrhea for <12 months prior to Screening, or not surgically sterile [absence of ovaries and/or uterus]), a serum pregnancy test will be performed at Screening and urine pregnancy tests will be performed prior to dosing at all other visits, as applicable, and at the End-of-Treatment Visit.
13. Patients who have decreased best-corrected visual acuity due to optic neuropathy (defined by a decrease in vision of 2 lines on the Snellen chart, new visual field defect, or color defect secondary to optic nerve involvement within the last 6 months) are not eligible for randomization.
14. Vital signs (blood pressure, heart rate, respiratory rate, temperature) will be measured at all clinic visits. Vital signs will be measured pre- and post-infusion on Day 1 and Week 3 (all patients), pre- and post-infusion at Week 24 and Week 27 (proptosis non-responders who elect to receive open-label TEPEZZA) and pre-infusion on all other infusion days. Additional vital signs will be monitored if infusion-associated AEs occur.
15. Patients must be euthyroid, with the Baseline disease under control or have mild hypo- or hyperthyroidism (defined as FT_4 and FT_3 levels <50% above or below the normal limits). Every effort should be made to correct the mild hypo- or hyperthyroidism promptly and to maintain the euthyroid state for the full duration of the trial.
16. HbA1c must be ≤8.0% to be eligible for the trial. If the HbA1c is elevated and considered clinically significant at any time point after Screening, it will be repeated approximately every 90 days until it returns to normal or Baseline value.
17. ADA sample will be collected prior to the infusion. If the ADA test is positive after confirmatory and reactive titer testing, the sample will then be tested for NAb. If the patient tests positive for NAb, the patient may be followed until levels either return to Baseline or the patient's level decreases or remains stable. ADA samples will be collected during any visit triggered by suspected immunologically related AEs.
18. Pharmacokinetic samples will be collected prior to, and at the end of, the infusion on Day 1, Week 3 and Week 12 of the double-masked Treatment Period (all patients) and Week 24, Week 27 and Week 36 of the open-label Treatment Period (proptosis non-responders who elect to receive open-label TEPEZZA); a single sample will be collected at Week 24 (proptosis responders and proptosis non-responders who elect not to receive open-label TEPEZZA) and Week 48 (proptosis non-responders who elect to receive open-label TEPEZZA).

19. CCI [redacted]
20. Not collected for patients who prematurely discontinue from the double-masked Treatment Period.
21. For those patients who have consented, photographs of the patient's eyes will be taken prior to the first infusion and at Week 24 (all patients) and Week 48 (non-responders who elect to receive open-label TEPEZZA).
22. For non-responders who choose to receive open-label treatment, TEPEZZA infusion will be performed at Week 24 (first of 8 infusions). All Week 24 assessments except AE and concomitant medication monitoring, post-infusion PK sample collection and 24-hour post-dose phone/email contact must be completed prior to TEPEZZA infusion.
23. Phone (or email) contact by research staff focusing on safety and tolerability aspects will be made the day after infusion for the first and second infusions during the double-masked Treatment Period and during the open-label Treatment Period, and thereafter as deemed appropriate. In addition, patients who experience an infusion-associated event after any subsequent infusion will also be contacted by phone (or email) by research staff the day after the infusion, and thereafter as deemed appropriate.

15.2 Unit Conversion

Quantity	From SI to Conventional Unit	From Conventional Unit to SI
Height	0.394 x [cm] = [in]	2.540 x [in]. = [cm]
Weight	2.205 x [kg] = [lb]	0.454 x [lb] = [kg]
Temperature	([°C] x 9/5) + 32 = [°F]	([°F] – 32) 5/9 = [°C]

15.3 Graves' Ophthalmopathy Quality of Life Questionnaire

Directions:

- The following questions deal specifically with your thyroid eye disease. **Please focus on the past week while answering these questions.**
- Please tick only one box that matches your answer. The boxes correspond with the answers above them.

15.4 CCI

Research Site Personnel Instructions:

Use a ruler to measure the distance, in mm, from the origin (0) to the patient's mark. That distance will be the patient's measure of pain intensity.

Distance measured left Eye: _____ mm
Distance measured right Eye: _____ mm

15.5 CCI

15.5.1 CCI

CCI

CCI

CCI

15.5.2 CCI

CCI

15.5.3 CCI

CCI


CCI

15.6 CCI

15.6.1 CCI

15.6.3 CCI

CCI