Statistical Analysis Plan

[redacted]

Document Type:	Template	Document ID:	03.007A.03
Issue Date:	02 MAY 2016	Effective Date:	30 MAY 2016

Sponsor Name:	Horizon Pharma USA, Inc. 150 S. Saunders Road Lake Forest, IL 60045
Protocol Number and Title:	HZNP-TEP-301: A Phase 3, Randomized, Double-Masked, Placebo-Controlled, Parallel-Group, Multicenter Study Evaluating Teprotumumab (HZN-001) Treatment in Subjects with Active Thyroid Eye Disease (OPTIC)
Protocol Version and Date:	Version 1.0: 10-Jul-2017 Version 1.1: 24-Jan-2018 Version 2.0: 24-Oct-2017 Version 3.0: 16-Apr-2018 Version 4.0: 31-Jan-2019
INC Research Project Code:	[redacted]
Author(s):	[redacted] Manager, Biostatistics
SAP Version:	1.0
SAP Version Date:	08-Feb-2019

In the event of an actual or suspected breach of this obligation, INC Research should be notified promptly.

Revision History

Version #	Date (dd-mmm-yyyy)	Document Owner	Revision Summary

I confirm that I have reviewed this document and agree with the content.

APPROVALS	
INC Research	
[redacted] Manager, Biostatistics	08-Feb 2019 **Date** (dd-Mmm-yyyy)
[redacted] Sr. Director, Biostatistics	08FEB2019 **Date** (dd-Mmm-yyyy)
Horizon Pharma USA, Inc.	
[redacted] Director, Clinical Development and Operations	**Date** (dd-Mmm-yyyy)
[redacted] Associate Director, Biostatistics	**Date** (dd-Mmm-yyyy)
[redacted] Senior Medical Director, Clinical Development	**Date** (dd-Mmm-yyyy)

TABLE OF CONTENTS

Contents

1. GLOSSARY OF ABBREVIATIONS

Abbreviation	Description
A	Appearance
ADA	Anti-drug Antibody
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
APD	Afferent Pupillary Defect
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
BLQ	Below the Limit of Quantification
CAS	Clinical Activity Score
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CS	Clinically Significant
CSH	Heterogeneous Compound Symmetry
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of Variation
DSMB	Data Safety Monitoring Board
EAIR	Exposure-Adjusted Incidence Rates
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EU	Europe
EUGOGO	European Group on Graves' Ophthalmopathy
FT3	Free Triiodothyronine
FT4	Free Thyroxine
GO-QoL	Graves' Ophthalmopathy Quality of Life
HLGT	High Level Group Term
HLT	High Level Term

Abbreviation	Description
ICD	Intercanthal Distance
ICH	International Conference on Harmonisation
IEC	Independent Ethics Committee
IGF-1R	Insulin-like Growth Factor-1 Receptor
IRB	Institutional Review Board
ITT	Intent-to-Treat
IV	Intravenous
IWRS	Interactive Web Response System
kg	Kilograms
lbs	Pounds
LLOQ	Lower Level of Quantification
LR	Light Reflex
LS	Least Squares
mAb	Monoclonal Antibody
Max	Maximum
MedDRA	Medical Dictionary for Regulatory Activities
Min	Minimum
mm	Millimeters
MMRM	Mixed Model Repeated Measures
mITT	Modified Intent-to-Treat
n	Number of Observations
NA	Not Applicable
NAb	Neutralizing Antibody
NaCl	Sodium Chloride
NCS	Not Clinically Significant
NE	Not Estimable
PD	Pharmacodynamic
PK	Pharmacokinetic
PP	Per Protocol

Abbreviation	Description
PT	Preferred Term
PW	Premature Withdrawal
q3W	Once Every Three Weeks
QoL	Quality of Life
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SE	Standard Error
SI	Standard International System of Units
SMQ	Standardised MedDRA Query
SOC	System Organ Class
SOP	Standard Operating Procedure
TEAE	Treatment-Emergent Adverse Event
TED	Thyroid Eye Disease
TLF	Table, Listing and Figure
TOEP	Toeplitz
TOEPH	Heterogeneous Toeplitz
ULN	Upper Limit of Normal
UN	Unstructured
US	United States
VF	Visual Functioning
WHO-DD	World Health Organization Drug Dictionary

2. PURPOSE

The purpose of this statistical analysis plan (SAP) is to ensure that the data listings, summary tables and figures which will be produced, and the statistical methodologies that will be used, are complete and appropriate to allow valid conclusions regarding the study objectives.

2.1. RESPONSIBILITIES

[redacted] will perform the statistical analyses and is responsible for the production and quality control of all tables, figures and listings. A separate vendor will be responsible for generating the pharmacokinetic analysis plan and associated analyses with the exception of listings and descriptive summaries of concentration data which will be performed by [redacted] as outlined in subsequent sections.

2.2. TIMINGS OF ANALYSES

Data Safety Monitoring Board (DSMB) Analysis: The study will be monitored by a DSMB, which will advise the Sponsor regarding the continuing safety of study subjects. The details regarding frequency of meetings, members, and the safety review criteria will be outlined in a separate DSMB Charter. An independent unmasked team from [redacted] biostatistics will perform the analyses to maintain the masking of the study.

Primary Study Analysis: The primary analysis of safety, efficacy, and pharmacokinetics is planned after all subjects complete the final visit (Week 24) in the double-masked Treatment Period or terminate early from the study. The analysis includes all data collected in the database through the time of the database lock for the Week 24 cut-off (adverse events and concomitant medications occurring up to and including 21 days after the last dose of study drug and all data collected at visits through Week 24).

Follow-up Analysis: A follow-up analysis will be conducted when all subjects complete the final visit (Week 72) of the Follow-Up Period of the study or terminate early. Additional data collected after the database lock from the primary analysis of the study will be prepared as an addendum to the Clinical Study Report according to regulatory or scientific need.

3. STUDY OBJECTIVES

3.1. PRIMARY OBJECTIVE

The overall objective is to investigate the efficacy, tolerability, and safety of teprotumumab (HZN-001, a fully human monoclonal antibody [mAb] inhibitor of the insulin-like growth factor-1 receptor [IGF-1R]) administered once every 3 weeks (q3W) for 24 weeks, in comparison to placebo, in the treatment of subjects with moderate-to-severe active thyroid eye disease (TED).

The primary objective is to evaluate the effect of teprotumumab versus placebo on the proptosis responder rate (i.e., the percentage of subjects with a ≥ 2 millimeter [mm] reduction from Baseline in the study eye without deterioration [≥ 2 mm increase] of proptosis in the fellow eye) at Week 24.

3.2. SECONDARY OBJECTIVES

Secondary objectives are (analyzed hierarchically):

1. Evaluate the effect of teprotumumab versus placebo on the overall responder rate (percentage of subjects with ≥ 2-point reduction in Clinical Activity Score [CAS] AND ≥ 2 mm reduction in proptosis from Baseline, provided there is no corresponding deterioration [≥ 2-point/mm increase] in CAS or proptosis in the fellow eye) at Week 24.
2. Evaluate the effect of teprotumumab versus placebo on the percentage of subjects with a CAS value of 0 or 1 at Week 24 in the study eye.
3. Evaluate the effect of teprotumumab versus placebo on the mean change from Baseline to Week 24 in proptosis measurement in the study eye.
4. Evaluate the effect of teprotumumab versus placebo on the improvement in diplopia (percentage of subjects with baseline diplopia >0 in study eye who have a reduction of ≥ 1 grade with no corresponding deterioration [≥1 point worsening] in the fellow eye) at Week 24.
5. Evaluate the effect of teprotumumab versus placebo on the mean change from Baseline to Week 24 in the Graves' Ophthalmopathy Quality of Life (GO-QoL) questionnaire overall score.

3.3. EXPLORATORY OBJECTIVES

Exploratory objectives are:

1. Evaluate the effect of teprotumumab versus placebo on the Clinical Measures of Severity individual response status frequencies and percentage of responders for each component of clinical severity from Baseline to Week 24.
2. Evaluate the effect of teprotumumab versus placebo on the mean change from Baseline to Week 24 in the CAS.
3. Evaluate the effect of teprotumumab versus placebo on the overall responder rate at Week 24 stratified by the level of response (high responders, responders, low responders, and non-responders; see Section 8.3.3 for definitions).
4. Evaluate pharmacokinetic (PK) parameters of teprotumumab to estimate exposure and understand PK-pharmacodynamic (PD) relationships.
5. Evaluate the effect of teprotumumab versus placebo on the mean change from Baseline to Week 24 in the GO-QoL questionnaire visual functioning (VF) and appearance (A) subscale scores.
6. Evaluate the effect of teprotumumab on the mean change from Baseline to Week 24 in blood and serum biomarkers.
7. Evaluate the effect of teprotumumab versus placebo on the mean change from Baseline to Week 24 on the motility component of the Clinical Measures of Severity.

3.4. BRIEF DESCRIPTION

This study will be conducted at up to 16 sites in the United States (US) and Europe (EU).

This is a randomized, double-masked, placebo-controlled, parallel-group, multicenter study. Subjects will be screened for the study within 2 to 6 weeks prior to the Baseline (Day 1) Visit. Approximately 76 subjects (38/group) who meet the study eligibility criteria will be randomized via a central study-level randomization on Day 1 in a 1:1 ratio (stratified by tobacco use status) to receive 8 infusions of teprotumumab (10 mg/kg for the first infusion and 20 mg/kg for the remaining 7 infusions) or placebo q3W. All subjects will enter a 24-week double-masked Treatment Period, during which study 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). All study drug dosing will be performed at the clinic under the supervision of clinic staff, and at any scheduled infusion, the infusion rate may be reduced or the dose may be interrupted or held based on decreased tolerability. On each dosing day, scheduled assessments (except for AE and concomitant medication use monitoring, which will be monitored throughout the clinic visit) will be completed prior to study drug dosing. After each of the first 2 infusions, subjects will be contacted by phone/email the following day and will return to the clinic 1 week after the infusion (Weeks 1 and 4) for safety and tolerability assessments; additional phone/email contacts and clinic visits may also be conducted for any subject experiencing an infusion-associated event.

At the end of the double-masked Treatment Period (Week 24), subjects who are proptosis non-responders (study eye has < 2 mm decrease in proptosis) will be eligible to enter an open-label extension study (HZNP-TEP-302) in which subjects will receive 8 infusions of teprotumumab (10 mg/kg for the first infusion and 20 mg/kg for the remaining infusions) in an open-label fashion. Discussion of this option with the subject and informed consent presentation for HZNP-TEP-302 should be discussed at the Week 24 Visit given enrollment in HZNP-TEP-302 must occur within 2 weeks of the Week 24 Visit.

At Week 24, proptosis responders, as well as non-responders who choose not to enroll in the open-label extension study (HZNP-TEP-302), will enter a 48-week Follow-Up Period, during which study drug will not be administered and clinic visits are scheduled for Weeks 28, 36, 48, 60, and 72 (Months 7, 9,12, 15, and 18). Subjects who are responders at Week 24 but who meet the criteria for re-treatment due to a relapse during the Follow-Up Period may enroll in the open-label extension study, HZNP-TEP-302.

The criteria to determine relapse are the following:

- Increase in proptosis of ≥ 2 mm in the study eye since Week 24, or
- An increase in CAS ≥ 2 points since Week 24 with an absolute CAS of ≥ 4 in the study eye following the Week 24 visit.
- In addition to one of the bullet points above, the Investigator should also consider the subject's symptomology to ensure a relapse has occurred (e.g., new onset of double vision).

Subjects who prematurely discontinue study drug dosing prior to Week 21 of the Treatment Period will return to the clinic and undergo the scheduled Week 24 assessments and will be encouraged to participate in the Follow-Up Period unless they initiated another intervention due to lack of efficacy. Subjects who enter the Follow-Up Period but prematurely discontinue study participation prior to 48 weeks following the double-masked Treatment Period will return for a final visit and undergo the scheduled Week 72 assessments prior to study discharge.

Subjects who complete the Week 72 Visit will be contacted 6 and 12 months later via phone or email by research staff to enquire if any treatment for TED has been received since last study contact. If yes, subject will be questioned regarding type of treatment and outcome/response.

3.5. DETERMINATION OF SAMPLE SIZE

In the prior study, TED01RV, a 51% difference (71% vs 20%) between teprotumumab (10 mg/kg for the first infusion followed by 20 mg/kg for the remaining 7 infusions) and placebo was observed at Week 24 in favor of teprotumumab in proptosis reduction of

2 mm or more. A sample size of 38 subjects per group provides 90% power at the 2-sided alpha 0.05 level to detect a difference of 39% between teprotumumab and placebo; the sample size has been adjusted to allow for a 16% discontinuation rate.

3.6. TREATMENT ASSIGNMENT & MASKING

Once eligibility has been confirmed, subjects will be randomized to treatment groups in a 1:1 ratio, stratified by tobacco use status (non-user, user) via a central study-level randomization schedule. A randomization schedule will be generated by [redacted] prior to shipment of any study drug to the clinical sites. On Day 1 of the double-masked Treatment Period, once all Baseline procedures other than administration of drug have been completed, the masked site personnel will use the interactive web response system (IWRS) to randomize the subject. The unmasked pharmacist or designee will then use the IWRS to obtain dosing information, and dispense the appropriate study drug.

The pharmacists or designees responsible for preparing the teprotumumab or placebo solutions for intravenous (IV) administration will not be masked to the identity of the study drug. Pharmacists/designees will provide study drug in infusion bags (fully diluted for administration) to study site personnel with appropriate masked labels. The subject, Investigator, and all other study site personnel will be masked to the treatment being administered.

The study mask should be broken only if the safety of a subject is at risk and the treatment plan depends on which medication he or she received. Unless the subject is at immediate risk, the Investigator must make diligent attempts to contact the Sponsor or Sponsor's designee before unmasking the subject's data. If a subject's data are unmasked without prior knowledge of the Sponsor, the Investigator must notify the Sponsor as soon as possible and no later than the next business day. All circumstances surrounding the event must be clearly documented.

The Sponsor's Pharmacovigilance department or designee will unmask the identity of the study medication for an unexpected, drug-related serious adverse event (SAE) for submission to health authorities and Institutional Review Board (IRB)/Independent Ethics Committee (IEC) according to applicable regulatory requirements. However, the results will not be shared with other Sponsor representatives or staff at study sites. Details of subjects who are unmasked during the study will be included in the Clinical Study Report.

Unmasking for independent pharmacological analysis of biological samples or SAE reporting will be performed according to procedures in place to ensure integrity of the data.

All investigative site staff directly involved in this study will remain masked from Screening through analysis of the Follow-Up data and all site close-out visits. Selected employees at the Sponsor and its designees will be unmasked after the database lock following completion of all subjects in the double-masked Treatment Period.

As PK, biomarker, and immunogenicity data have the potential to unmask treatment assignments, these data will not be analyzed by any masked staff until after the database lock following completion of all subjects in the double-masked Treatment Period.

An unmasked team from [redacted] biostatistics will perform unmasked analyses as outlined in the DSMB charter. These team members will otherwise have no involvement with the study.

3.7. PREPARATION AND ADMINISTRATION OF STUDY MEDICATION

During the double-masked Treatment Period, subjects will receive infusions of either:

1. Teprotumumab 20 mg/kg (10 mg/kg on Day 1 followed by 20 mg/kg q3W for the remaining 7 infusions), or
2. Placebo (q3W for all 8 infusions)

Teprotumumab will be provided in single-dose 20 mL glass vials as a freeze-dried powder. Each vial of teprotumumab will be reconstituted with 10 mL of water for infusion. The resulting solution will have an approximate concentration of 50 mg/mL teprotumumab. Reconstituted teprotumumab solution will be further diluted in 0.9% (w/v) sodium chloride (NaCl) solution prior to administration.

Doses up to 1800 mg will be administered in a total infusion volume of 100 mL, and those above 1800 mg will be administered in a total infusion volume of 250 mL. To maintain a constant volume in the infusion bags, a volume equal to the volume of teprotumumab to be placed into the infusion bag will be first removed from the infusion bag using a sterile syringe and needle. The appropriate volume of reconstituted drug product solution based on the subject`s dose and body weight will be withdrawn and the teprotumumab reconstituted drug product solution will be diluted with normal saline (0.9% NaCl) in the infusion bag.

Placebo will consist of a normal saline (0.9% NaCl) solution and will be administered in 100 mL or 250 mL infusion bags, as appropriate, per weight-based dosing volumes.

The infusion rate may be reduced and the dose may be interrupted or held based on tolerability. The first and second infusions will be administered over approximately 90 minutes (but not less than 80 minutes). Subsequent infusions will be administered over approximately 60 minutes (but not less than 50 minutes) providing there are no significant infusion-associated events.

3.8. STUDY PROCEDURES AND FLOWCHART

The Schedule of Assessments is presented in Table 1 below.

Table 1: Schedule of Assessments

	Screening [1]	Treatment Period [2]											Follow-Up Period [3]					Follow-Up Contact [4]	
Study Visit	S1/S2/S3	1	2	3	4	5	6	7	8	9	10	11/ PW1 [5]	12	13	14	15	16/ PW2 [6]	17	18
Week (W)/ Month (M)	-42 to -14 days	Day 1 [7]	W1	W3	W4/ M1	W6	W9	W12/ M3	W15	W18	W21	W24/ M6	W28/ M7	W36/ M9	W48/ M12	W60/ M15	W72/ M18	W96/ M42	W120/ M66
Visit Window (± days)		(±3)	(±1)	(±3)	(±1)	(±3)	(±3)	(±3)	(±3)	(±3)	(±3)	(±7)	(±7)	(±7)	(±7)	(±7)	(±7)	(±14)	(±14)
Informed Consent	X																		
Review inc/exc criteria	X	X																	
Demographics	X																		
Medical History [8]	X [9]	X																	
Weight [10]	X							X				X		X	X	X	X		
Randomization [11]		X [7]																	
Study drug infusion		X		X		X	X	X	X	X	X								
Phone (email) contact for safety 24 hours postdose [12]		X		X															
Efficacy assessments																			
CAS [13]	X	X [14]				X		X		X		X	X	X	X	X	X		
Clinical Measures of Severity - includes proptosis, diplopia and motility restriction	X	X [15]				X		X		X		X	X	X	X	X	X		
Safety assessments																			
Pregnancy Test [16]	X	X		X		X	X	X	X	X	X	X	X	X	X		X [17]		

	Screening [1]	Treatment Period [2]											Follow-Up Period [3]					Follow-Up Contact [4]	
Study Visit	S1/S2/S3	1	2	3	4	5	6	7	8	9	10	11/ PW1 [5]	12	13	14	15	16/ PW2 [6]	17	18
Week (W)/ Month (M)	-42 to -14 days	Day 1 [7]	W1	W3	W4/ M1	W6	W9	W12/ M3	W15	W18	W21	W24/ M6	W28/ M7	W36/ M9	W48/ M12	W60/ M15	W72/ M18	W96/ M42	W120/ M66
Visit Window (± days)		(±3)	(±1)	(±3)	(±1)	(±3)	(±3)	(±3)	(±3)	(±3)	(±3)	(±7)	(±7)	(±7)	(±7)	(±7)	(±7)	(±14)	(±14)
Physical exam [18]	X [19]	X [18]	X			X		X		X		X [18]			X		X [18]		
Ophthalmic exam [20]	X [21]	X	X			X		X		X		X			X		X		
Vital Signs [22]	X	X [22]	X	X [22]	X	X	X	X	X	X	X	X	X	X	X	X	X		
12-Lead ECG	X	X		X		X		X				X					X		
Clinical laboratory tests [23]																			
Chemistry (excl. glucose)	X [24]	X		X		X	X	X		X		X		X			X		
Thyroid (FT3, FT4, TSH) [25]	X	X		X		X	X	X		X		X		X			X		
Hematology	X	X	X	X	X	X	X	X	X	X	X	X		X			X		
Glucose [23]	X	X	X	X	X	X	X	X	X	X	X	X		X			X		
HbA1c [26]	X							X				X		X			X		
Urinalysis	X	X		X		X	X	X		X		X		X			X		
ADA/NAb samples [27]		X		X			X					X [28]		X			X		
AE, SAE assessment [29]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
Concomitant medications	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
GO-QoL Questionnaire		X				X		X				X	X		X		X		
Biomarker samples [30]		X						X				X							
PK samples [31]		X	X	X	X		X					X [28]							
Contact (phone/email) to																		X	X

	Screening [1]	Treatment Period [2]											Follow-Up Period [3]					Follow-Up Contact [4]	
Study Visit	S1/S2/S3	1	2	3	4	5	6	7	8	9	10	11/ PW1 [5]	12	13	14	15	16/ PW2 [6]	17	18
Week (W)/ Month (M)	-42 to -14 days	Day 1 [7]	W1	W3	W4/ M1	W6	W9	W12/ M3	W15	W18	W21	W24/ M6	W28/ M7	W36/ M9	W48/ M12	W60/ M15	W72/ M18	W96/ M42	W120/ M66
Visit Window (± days)		(±3)	(±1)	(±3)	(±1)	(±3)	(±3)	(±3)	(±3)	(±3)	(±3)	(±7)	(±7)	(±7)	(±7)	(±7)	(±7)	(±14)	(±14)
assess additional TED treatment [32]																			

ADA=anti-drug antibody; AE=adverse event; CAS=Clinical Activity Score; ECG=electrocardiogram; FT3=free triiodothyronine; FT4=free thyroxine; FU=Follow-Up; GO-QoL=Graves' Ophthalmopathy Quality of Life Questionnaire; HbA1c=glycated hemoglobin; M=month; NAb=neutralizing antibody; PK=pharmacokinetic; PW=premature withdrawal; SAE=serious adverse event; TED=thyroid eye disease; TSH=thyroid stimulating hormone; W=week.

Footnotes:

1. Screening procedures can take place over more than 1 day/clinic visit provided consent is obtained first and all assessments are completed within the designated window.
2. Double-masked Treatment Period. Subjects who are proptosis non-responders at Week 24 are eligible to enroll in an open-label extension study in which all subjects will receive teprotumumab 20 mg/kg (10 mg/kg for the first infusion and 20 mg/kg for the remaining infusions).
3. Proptosis responders and non-responders who choose not to enroll in the open-label extension study will participate in a 48-week Follow-Up Period.
4. Subjects who complete the Week 72 Visit will be contacted via phone or email by research staff to enquire if any treatment for TED has been received since last study contact.
5. If a subject prematurely discontinues study drug during the Treatment Period, they will return for a clinic visit and undergo the Week 24 assessments, with the exception of the collection of blood samples for PK and ADA evaluations. Subjects will be encouraged to continue study participation in the Follow-Up Period.
6. If a subject prematurely discontinues from the study during the Follow-Up Period, they will return for a clinic visit and undergo the Week 72 assessments prior to discharge.
7. On Day 1 (Baseline), subjects will be randomized and receive the first dose of study drug; however, Baseline assessments will be performed prior to dosing.
8. Medical history including tobacco use history and Graves' disease and treatment history.
9. TED must be moderate to severe in intensity (non-sight threatening but appreciable impact on daily life) with an onset of symptoms (as determined by subject records) within 9 months prior to the Baseline Visit for study enrollment.
10. Dosing will be adjusted if there is a change in weight during the Treatment Period. The weight obtained at Week 12 can be used in dose calculations beginning at Week 12 or Week 15.

11. Subjects will be randomized in a 1:1 ratio (stratified by tobacco use status) to receive either: a) teprotumumab (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).
12. 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, and thereafter as deemed appropriate. In addition, subjects 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.
13. CAS must be ≥ 4 for enrollment and randomization.
14. Subjects whose CAS score decreases 2 or more points in the study eye from Screening are not eligible for randomization.
15. Subjects who have a ≥ 2 mm decrease in proptosis in the study eye from Screening are not eligible for randomization.
16. Serum pregnancy test at Screening and urine pregnancy tests prior to dose at all other visits, as applicable. Perform for female subjects 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]).
17. Pregnancy test only performed for female subjects of childbearing potential who enter the Follow-Up Period but discontinue study participation prior to Week 48.
18. Physical exam will include assessment of presence or absence of pretibial myxedema on Day 1 and Week 24 (or PW) of the Treatment Period and Week 72 (or PW) of the Follow-Up Period. If present, measurements of instep and calf will be taken.
19. Height will be measured at Screening only.
20. Ophthalmic exam: best corrected visual acuity, pupil exam, color vision assessment, Ishihara color plates (or equivalent) or related red desaturation, intraocular pressure, and slit lamp exam. If significant abnormalities are noted compared to previous visits, including a loss of 2 lines or more of vision, development of pupil abnormalities including APD, rise in intraocular pressure, development of corneal infiltrates or other abnormalities not here specified but of concern to the ophthalmologist, further investigations of visual function will be conducted according to the ophthalmologist decision.
21. Subjects 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.
22. Vital signs (heart rate, blood pressure, 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, and pre-dose on all other infusion days. Additional vital signs will be monitored if infusion-associated AEs occur.
23. Non-diabetic subjects should be fasting at Weeks 1 and 4 only. Diabetic subjects should be fasting at each visit blood glucose is evaluated.
24. ALT/AST must be ≤3 x the upper limit of normal (ULN) and serum creatinine must be <1.5 x the ULN according to age to be eligible for randomization.
25. Subjects must be euthyroid with the baseline disease under control or have mild hypo- or hyperthyroidism (defined as free thyroxine [FT4] and free triiodothyronine [FT3] 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 clinical trial.
26. HbA1c must be < 9.0% for randomization. If the HbA1c is elevated and considered clinically significant at any time point after Screening, it will be repeated approximately every 45 days until it returns to normal or baseline value.
27. If a sample is positive in the ADA test, after confirmatory and reactive titer testing, the sample will then be tested for NAb. If the subject tests positive for NAb, he/she will be followed until levels either revert to Baseline or the subject's value decreases or remains stable. Any subject with a positive NAb test at Week 72 (or PW) during the Follow-Up Period will continue to be followed until the subject's value decreases or remains stable.
28. Not collected for subjects who prematurely discontinue from the Treatment Period.

29. Adverse events (AEs) that occur within 2 weeks prior to Day 1 and prior to dosing on Day 1 will be considered baseline signs/symptoms. Adverse events occurring or worsening after the dose on Day 1 through the end of the Treatment Period will be considered treatment-emergent AEs (TEAEs). Adverse events occurring or worsening during the Follow-Up Period will be considered postdose AEs. All SAEs that occur from the signing of informed consent through 30 days after study discontinuation will be recorded.
30. Serum (two 5.0 mL samples) will be obtained on Day 1 and Weeks 12 and 24 of the Treatment Period for possible analysis of interleukin (IL)-4, IL-6, IL-10, IL-12, IL-13, IL-17, IL-23, IL-1ß, sIL-1RA, INFγ, TGFß, TNFα, micRNA and TSH-R-Ab. Based on the results of the above assays, other similar serum biomarkers may be assayed to further explore drug and disease mechanisms.
31. PK samples will be collected prior to, and at the end of, the infusion on Day 1 and Weeks 3 and 9 of the Treatment Period; additional single samples will be collected at Weeks 1, 4, and 24.
32. If TED treatment has been received since last contact, the subject will be questioned regarding type of treatment and outcome/response.

4. ENDPOINTS

4.1. PRIMARY EFFICACY ENDPOINT

The primary efficacy endpoint is the proptosis responder rate (percentage of subjects 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.

4.2. SECONDARY EFFICACY ENDPOINTS

The secondary endpoints are the following, to be analyzed in a hierarchical manner:

1. The overall responder rate (percentage of subjects with ≥ 2-point reduction in CAS AND ≥ 2 mm reduction in proptosis from Baseline, provided there is no corresponding deterioration [≥ 2 point/mm increase] in CAS or proptosis in the fellow eye) at Week 24.
2. Percentage of subjects with a CAS value of 0 or 1 in the study eye at Week 24.
3. Mean change from Baseline to Week 24 in proptosis measurement in the study eye.
4. Diplopia responder rate (percentage of subjects with baseline diplopia >0 in study eye who have a reduction of ≥ 1 grade with no corresponding deterioration [≥1 grade worsening] in the fellow eye) at Week 24.
5. Mean change from Baseline to Week 24 in the GO-QoL questionnaire overall score.

4.3. EXPLORATORY EFFICACY ENDPOINTS

Exploratory efficacy endpoints include:

1. Clinical Measures of Severity individual response status frequencies and percentage of responders for each component of clinical severity from Baseline to Week 24.
2. Mean change from Baseline to Week 24 in the CAS.
3. Overall responder rate at Week 24 stratified by the level of response (high responders, responders, low responders, and non-responders; see Section 8.3.3 for definitions).
4. Mean change from Baseline to Week 24 in the GO-QoL questionnaire VF and A subscale scores.
5. Mean change from Baseline to Week 24 on the motility component of the Clinical Measures of Severity.
6. Mean percent change from Baseline to Week 24 in proptosis measurement.
7. Time to relapse.

4.4. EXPLORATORY PHARMACOKINETIC AND BIOMARKER ENDPOINTS

The PK endpoints covered in the scope of this SAP includes:

- Descriptive summaries of serum concentrations by time point
- Listing of serum concentrations by time point

Biomarker endpoints include:

- Absolute concentrations for biomarkers (Section 10) at Day 1 and Weeks 12 and 24
- Change from Baseline for biomarkers (prior to dosing at Day 1 as the Baseline) at Weeks 12 and 24

4.5. SAFETY ENDPOINTS

Safety endpoints include the following:

- Adverse events (AE), including AE of special interest (AESI)
- Concomitant medication use monitoring
- Descriptive summary of immunogenicity
- Physical examination
- Ophthalmic examination
- Vital signs
- Clinical laboratory assessments (complete blood count, chemistry [including thyroid panel and HbA1C], and urinalysis)
- Pregnancy testing
- Electrocardiogram (ECG)

5. ANALYSIS POPULATIONS

5.1. INTENT-TO-TREAT POPULATION

The Intent-to-Treat (ITT) Population will include all subjects who are randomized to treatment. Subjects will be analyzed according to their randomized treatment assignment. The ITT Population will be used for all analyses of efficacy endpoints.

5.2. MODIFIED INTENT-TO-TREAT POPULATION

The Modified Intent-to-Treat (mITT) Population will include all ITT subjects who receive at least one dose of study drug and have at least one post-baseline measurement of primary efficacy. The mITT Population will be analyzed to evaluate the sensitivity of the primary efficacy analysis. Subjects will be analyzed according to their randomized treatment assignment.

5.3. SAFETY POPULATION

The Safety Population will include all subjects who receive at least one dose of study drug. Subjects will be analyzed according to treatment received. In the event a subject receives more than one treatment, the subject will be summarized by the treatment received most frequently. The Safety Population will be used for all analyses of safety endpoints.

5.4. PER PROTOCOL POPULATION

The Per Protocol (PP) Population includes all mITT subjects who complete the treatment period and do not incur any major protocol violations that would challenge the validity of their data. The PP Population will be analyzed to evaluate the sensitivity of the primary efficacy analysis and secondary efficacy analyses. Subjects will be analyzed according to treatment received. In the event a subject receives more than one treatment, the subject will be summarized by the treatment received most frequently. Major protocol violations that challenge the validity of the data will be identified by a masked review prior to the unmasking of subject data by Horizon.

5.5. PHARMACOKINETIC POPULATION

The PK Population will include all subjects who received at least one dose of study drug and have at least one quantifiable PK serum concentration measurement. Subjects will be analyzed according to treatment received.

PK serum concentrations of teprotumumab will be summarized by timepoint and listed for all subjects in the PK Population.

5.6. PROTOCOL DEVIATIONS

Protocol deviations will be categorized as 1) major vs. minor and 2) exclusionary vs non-exclusionary from PP Population, prior to the unmasking of the treatment groups at the end of the double-masked Treatment Period. Any deviations that accumulate during the Follow-Up Period will be subsequently categorized as major vs. minor prior to the final database lock but will not be assessed for exclusionary purposes for the PP Population. Major protocol deviations will be summarized for the ITT Population by the number and percentage of subjects overall, by treatment group, and category of deviation during the double-masked Treatment Period and Follow-Up Period, separately. All protocol deviations will be listed.

6. GENERAL ASPECTS FOR STATISTICAL ANALYSIS

6.1. GENERAL METHODS

The following conventions will be utilized in the analyses:

- In general, descriptive summaries will be provided by treatment group and by visit (double-masked Treatment Period and Follow-Up Period visits as appropriate).
- For by-visit descriptive summaries presented for the double-masked Treatment Period, summaries will also be included for Follow-Up Period visits for those subjects who did not continue onto the open-label extension study (HZNP-TEP-302).
- The following treatment group labels will be used: "Placebo" and "HZN-001" for teprotumumab 10 mg/kg first infusion and teprotumumab 20 mg/kg remaining 7 infusions.
- Unless otherwise indicated, continuous variables will be summarized using the number of observations (n), mean, standard deviation (SD), median, minimum (min), and maximum (max). Categorical variables will be summarized using number of observations (n), frequency and percentages of subjects.
- The same number of decimal places as the raw data will be presented when reporting min and max, 1 more decimal place than in the raw data will be presented when reporting mean and median, and 2 more decimal places than in the raw data will be presented when reporting SD.
- All statistical tests will be two-sided and compared to the 5% significance level. Confidence intervals (CI) will be based on 95% confidence and two-sided.
- If multiple assessments occur at a given time point, the latest value will be used. If multiple assessments occur on the same day and the time is not recorded, the average of the values will be used in any descriptive summaries.
- All relevant subject data will be included in listings. All subjects entered into the database will be included in subject data listings.
- Additional programming considerations are provided in Section 15.

6.2. KEY DEFINITIONS

6.2.1. Study Eye

The most severely affected eye will be defined as the "study eye" at the Baseline (Day 1) visit. If there is a discrepancy between CAS and proptosis in determining the study eye, this will be adjudicated always to the eye with the most significant proptosis. 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 outcome measure.

6.2.2. Fellow Eye

The non-study eye will be referred to as the “fellow eye”.

6.2.3. Baseline

Baseline will be defined as the last measurement taken prior to first dosing (considering unscheduled visits when available). Change from Baseline will be defined as the measurement at each time point minus the Baseline value.

6.2.4. Study Day

For study days on or after the date of treatment start, Study Day will be calculated as assessment date – first dose date + 1. For study days prior to dosing, Study Day will be calculated as assessment date – treatment start date. Further, there will be no Study Day 0.

6.3. MISSING DATA

6.3.1. 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.

6.3.2. 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.

6.3.3. 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 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 month is 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 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 both prior and concomitant. If the stop date is prior to the first dose date, the medication will be considered to be prior only. If the stop date is after the first dose date but prior to the end of the Week 24 visit window (Week 24 + 1 week), the medication will be considered prior and concomitant during the double-masked Treatment Period only. If the stop date is after the Week 24 visit window, the medication will be considered prior and concomitant during both the double-masked Treatment Period and Follow-Up Period.
- 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 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 date of last study visit will be used.

6.3.4. 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 if an AE is treatment-emergent. Imputed dates will not appear in the data listings.

For partial 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 month is 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 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 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.

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.

6.3.5. Efficacy Assessments

For the primary analysis of binary primary endpoint and secondary endpoints, subjects missing the Week 24 evaluation will be treated as failures (non-responders). Sensitivity analyses of the primary endpoint are presented in Section 8.1.2 to assess the impact of missing data.

No imputation of missing data will be performed for the secondary analysis of proptosis and GO-QoL as continuous variables, as the analysis will be performed using Mixed Model Repeated Measures (MMRM) models, which accommodate missing data and allow information from every visit to contribute to the analysis of the data. The only exception to this is that for any subject in the ITT Population 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 from the MMRM analysis).

6.4. VISIT WINDOWS

All data will be summarized according to the scheduled visit and time points as outlined in the protocol and by the visit denoted on the electronic case report form (eCRF). Further, premature withdrawal (PW) visits will be windowed to the nearest scheduled visit based on the study day of occurrence relative to the target day of each scheduled visit according to Table 1 below.

Table 1: Visit Windows for Assigning PW Visits to a Scheduled Visit

Study Period	Visit	Target Day	Window for Reassignment to Scheduled Visit
Double-Masked Treatment Period	Day 1	1	1
	Week 1	8	2 - 15
	Week 3	22	16 - 25
	Week 4 (Month 1)	29	26 - 36
	Week 6	43	37 - 53
	Week 9	64	54 - 74
	Week 12 (Month 3)	85	75 - 95
	Week 15	106	96 - 116
	Week 18	127	117 - 137
	Week 21	148	138 - 158
	Week 24 (Month 6)	169	159 - 176
Follow-Up Period	Week 28 (Month 7)	197	177 - 225
	Week 36 (Month 9)	253	226 - 295
	Week 48 (Month 12)	337	296 - 379
	Week 60 (Month 15)	421	380 - 463
	Week 72 (Month 18)	505	≥ 464

In the event that a PW visit gets reassigned to a visit for which the subject has scheduled data collected, the data from the nominal scheduled visit will take precedence and the data from the PW visit will not be summarized.

6.5. POOLING OF SITES

In general, data from all sites will be summarized together for efficacy and safety analyses. Select descriptive summaries will be provided by site and region: US and EU.

6.6. SUBGROUPS

Subgroups of interest include:

- Tobacco use status (non-user, user): Tobacco uses of "never" and "former" will be grouped as tobacco non-users, and tobacco uses of "current" will be considered tobacco users as collected on the substance use eCRF.
- Region (US, EU)
- Site
- Age (<65 years, ≥65 years)
- Gender (male, female)

- Race (white, black or African American, Asian, other)
- Ethnicity (Hispanic or Latino, not Hispanic or Latino)

Descriptive summaries will be provided for various efficacy endpoints and TEAEs for select subgroups as specified in Section 8 and Section 11.3, respectively. If one of the subgroups has less than 5 subjects, no summary table will be created, except for site and race. Sites will not be pooled regardless of subject counts in the by-site subgroup analyses.

6.7. MULTIPLICITY

If a statistically significant result in favor of teprotumumab is achieved on the primary endpoint, the secondary endpoints (with the main analyses defined in the study eye as appropriate) will be analyzed in a hierarchical manner to control for multiplicity. For each secondary outcome measure, teprotumumab will be tested against placebo at the 0.05 significance level only if the test statistic was statistically significantly in favor of teprotumumab for the outcome measure preceding it in the following hierarchical order:

1. The overall responder rate (percentage of subjects with ≥ 2-point reduction in CAS AND ≥ 2 mm reduction in proptosis from Baseline, provided there is no corresponding deterioration [≥ 2 point/mm increase] in CAS or proptosis in the fellow eye) at Week 24.
2. Percentage of subjects with a CAS value of 0 or 1 in the study eye at Week 24.
3. Mean change from Baseline to Week 24 in proptosis measurement in the study eye.
4. Diplopia responder rate (percentage of subjects with baseline diplopia >0 in the study eye who have a reduction of ≥ 1 grade with no corresponding deterioration [≥1 grade worsening] in the fellow eye) at Week 24.
5. Mean change from Baseline to Week 24 in the GO-QoL questionnaire overall score.

7. DEMOGRAPHIC, OTHER BASELINE CHARACTERISTICS AND MEDICATION

7.1. SUBJECT DISPOSITION AND DISCONTINUATION

A summary of subject disposition will be provided including the number of subjects screened and number of screen failures, as well as the number of subjects in each analysis population (ITT, mITT, Safety, PP, and PK Populations). In addition, the number and percent of subjects in each of the following categories will be provided:

- Completed the double-masked Treatment Period
- Discontinued early from the double-masked Treatment Period overall
    - Discontinued early from the double-masked Treatment Period and did not continue in the Follow-Up Period
    - Discontinued early from the double-masked Treatment Period and continued in the Follow-Up Period
    - Reasons for early discontinuation from the double-masked Treatment Period
- Enrolled in the open-label extension study overall
    - Last visit completed
- Had any data collected during the Follow-Up Period
- Met relapse criteria during the Follow-Up Period overall and by reason for relapse
- Completed the study overall
    - Completed the double-masked Treatment Period and enrolled in the open-label extension study
    - Completed the Follow-Up Period for subjects who did not enroll in the open-label extension study
        - Completed the double-masked Treatment Period
        - Discontinued from the double-masked Treatment Period
- Discontinued early from the study overall
    - Did not complete the double-masked Treatment Period and did not continue in the Follow-Up Period
    - Did not complete the double-masked Treatment Period, continued in the Follow-Up Period but did not complete the Follow-Up Period

- Completed the double-masked Treatment Period but did not continue in the Follow-Up Period for subjects who did not enroll in the open-label extension study
- Completed the double-masked Treatment Period, continued in the Follow-Up Period but did not complete the Follow-Up Period for subjects who did not enroll in the open-label extension study
- Completed the double-masked Treatment Period and relapsed in the Follow-Up Period and enrolled in the open-label extension study
- Reasons for early study discontinuation during the Follow-Up Period

- Reasons for early study discontinuation for all subjects

Percentages will be based on the number of subjects in the ITT Population.

Subject disposition information will be provided overall and by treatment group.

A separate summary will be provided of the number and percentage of subjects attending each visit. Percentages will be based on the ITT Population.

7.2. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Descriptive summaries of demographic and baseline characteristics will be presented overall and by treatment group for the ITT Population and Safety Population. These characteristics include study eye (right, left), age, age category (<65 years, ≥65 years), gender, race, ethnicity, height, weight, body mass index, region (US, EU), tobacco use status as randomized (non-user, user), actual tobacco use status as collected on the substance use CRF (non-user, user), tobacco use history (never, current, former), alcohol use history (never, current, former), child-bearing potential (yes, no, not applicable), time since diagnosis of Graves' disease and time since diagnosis of TED. Further, free triiodothyronine (FT3), free thyroxine (FT4), thyroid stimulating hormone (TSH) and glycated hemoglobin A1C (HbA1C) levels will be summarized.

Age will be calculated as: (informed consent date - date of birth + 1) / 365.25 and truncated to complete years. Missing data imputation rules are provided in Section 6.3.1.

Weight will be converted to kilograms (kg) when reported in pounds (lbs) as follows:

Weight (in kg) = weight (in lbs) * 0.4536

Time since Graves' disease diagnosis (years) will be calculated as: (first dose date - date of diagnosis + 1) / 365.25, rounded to two decimal places. Time since TED disease diagnosis (months) will be calculated as: (first dose date - date of diagnosis + 1) / 30.4, rounded to two decimal places. Missing data imputation rules are provided in Section 6.3.2.

Separate summaries will be provided by tobacco use (non-user, user) and region (US, EU) for the ITT Population. Demographic data and baseline characteristics will be provided in subject listings.

7.3. MEDICAL HISTORY

Medical history information will be coded using Medical Dictionary for Regulatory Activities (MedDRA) version 20.1, summarized and presented overall and by treatment group based on the Safety Population. Summaries will be ordered alphabetically by system organ class (SOC) and then, within a SOC, alphabetically by preferred term (PT).

7.4. MEDICATION

Medications will be coded using World Health Organization Drug Dictionary (WHO-DD) September 2017. Prior and concomitant medications will be summarized by presenting the counts and percentage of subjects using medications overall and by each treatment group for the Safety Population. Summaries will be provided by Anatomical Therapeutic Chemical (ATC) Level 4 term and PT. Medication summaries will be sorted alphabetically by ATC Level 4 and by PT within ATC Level 4. Subjects will be counted only once for each medication class and each preferred drug name.

Prior and concomitant medications will be listed together with a designation to identify the medications as prior and/or concomitant and sorted by start date.

7.4.1. Prior Medication

Prior medications will be presented separately from concomitant medications in a summary table. Any medication with a start date prior to the date of first dose will be considered a prior medication. Missing date imputation rules are provided in Section 6.3.3.

7.4.2. Concomitant Medication

Concomitant medications will be presented in two separate summary tables: concomitant medications occurring during the double-masked Treatment Period (up to and including 21 days after the last dose of study drug) and concomitant medications occurring during the Follow-Up Period (the period after 21 days following the last dose of study drug). Any medication that is ongoing or has a stop date on or after the first dose date will be considered a concomitant medication. Any concomitant medication that has a start date, stop date or is ongoing on or after the first dose but prior to 21 days following the last dose of study drug will be considered a concomitant medication during the double-masked Treatment Period. Any medication that is ongoing or has a stop date after 21 days following the last dose of study drug will be considered a concomitant medication during the Follow-Up Period. As such, the same medication may be summarized as both prior and concomitant during either or both periods. The summary tables will not be mutually exclusive. Missing date imputation rules are provided in Section 6.3.3.

7.4.3. Concomitant Procedures

A listing will be provided for subjects undergoing any concomitant procedures.

8. EFFICACY

Efficacy assessments will be performed for both eyes at each assessment time point; however, only the study eye will be used for the primary endpoint and the analyses for secondary endpoints as applicable (Section 4.2).

8.1. PRIMARY EFFICACY ENDPOINT AND ANALYSIS

The primary efficacy endpoint is the proptosis responder rate (percentage of subjects 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 in the ITT Population.

Proptosis assessments will be performed using a Hertel exophthalmometer provided by the Sponsor for consistency in measurement, and (except when strictly unavoidable) the same Hertel instrument and same observer should be used at each evaluation for the full duration of the study. Additionally, the same intercanthal distance (ICD) must be used on each occasion.

Proptosis will be measured for each eye at Screening, Day 1 and Weeks 6, 12, 18, and 24 (or PW) during the Treatment Period, and at Weeks 28, 36, 48, 60 and 72 (or PW) during the Follow-Up Period). Subjects who have a ≥ 2 mm decrease in proptosis from Screening to Baseline in the study eye are not eligible for randomization. Measurements will be recorded on the Clinical Measures of Severity eCRF under exophthalmos. The Baseline value will also be recorded.

8.1.1. Primary Analysis of the Primary Endpoint

The analysis of the primary proptosis responder endpoint will assess the stratified difference in the proportions of proptosis responders between the treatment groups. Stratification for the analysis will use the same factor as was used to stratify randomization, tobacco use (non-user, user), with tobacco uses of “never” and “former” grouped as tobacco non-users, and tobacco uses of “current” considered tobacco users, as collected on the substance use eCRF. The primary analysis will use the ITT Population. The primary analysis will be a stratified difference, which is a weighted average of the difference within each stratum. Estimates from the two strata will be combined using Cochran-Mantel-Haenszel (CMH) weights. The test statistic will be calculated by dividing the stratified difference by the standard error. A two-sided p-value will be calculated assuming that the test statistic is distributed as a standard normal random variable under the null hypothesis. The analysis incorporating stratification is appropriate due to the randomization process incorporating stratification, as the analysis should follow randomization. The randomization was stratified by tobacco use due to the association between theoretical effect that tobacco

use and status can have on thyroid eye disease, further indicating that analysis incorporating tobacco use status is appropriate.

Subjects missing the Week 24 evaluation will be considered to be treatment failures (non-responders) for the primary analysis. Further, subjects who prematurely discontinue study drug dosing prior to Week 21 during the double-masked Treatment Period will be analyzed as treatment failures (non-responders), unless an assessment at Week 24 (within the day range specified in Section 6.4) is available.

The null hypothesis tested for the primary efficacy endpoint is that teprotumumab and placebo achieve equal proportions of proptosis responders at Week 24. The corresponding alternative hypothesis is that teprotumumab achieves an unequal proportion of proptosis responders at Week 24 compared to placebo. The primary efficacy endpoint will be achieved if the null hypothesis is rejected at the 0.05 alpha level in favor of teprotumumab.

The difference in response rates, comparing teprotumumab to placebo, will be estimated along with the corresponding 95% CIs and p-values.

Example SAS code is provided in the Appendix in Section 21.1.

8.1.2. Sensitivity Analysis of the Primary Endpoint

The primary efficacy analysis will be repeated for the mITT Population and the PP Population as sensitivity analyses. Additionally, the primary analysis will be repeated for the ITT Population using logistic regression, with treatment group as the model effect and tobacco use as a covariate in the model.

In order to evaluate the impact of missing data, the following additional sensitivity analyses will be conducted using the primary analysis method:

- Subjects missing the Week 24 evaluation will be analyzed using their last available assessment for classification of responder or non-responder for the ITT Population. Data collected from PW visits will be considered for this analysis.
- An analysis will be performed in which only subjects with a non-missing Week 24 evaluation will be included, regardless if they completed all scheduled treatments.

A chi-square test will be performed to evaluate the difference in percentage of responders between treatment groups at Week 24 for the ITT population, considering subjects missing the Week 24 evaluation as treatment failures (non-responders). The number and percentage of responders will be presented by treatment, along with the treatment difference, the corresponding asymptotic 95% CI and p-value.

A bar graph will be provided of the proportion of responders for the study eye by visit and treatment group for the ITT Population.

8.1.3. Subgroup Analysis of the Primary Endpoint

The primary efficacy analysis will be repeated separately for each tobacco use status (non-user, user) using an unstratified chi-square test within each status. A difference in response rates with a two-sided 95% confidence interval, calculated using normal theory, will be reported.

8.1.4. Additional Analysis of the Primary Endpoint

Further, the definition of responder will be applied for each study visit (Weeks 6, 12, and 18) and the analysis of risk difference will be repeated, considering subjects 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, a calculation of fellow eye response will be conducted (using a definition analogous to the one used for the study eye) at Week 24 and all other study visits (Weeks 6, 12, and 18). The primary analysis of risk difference will be conducted at Week 24 and all other study visits (Weeks 6, 12, and 18), considering subjects missing the evaluations as treatment failures (non-responders) and separately considering the observed results only.

Descriptive summaries of observed responder results will be provided by treatment group for each visit in the double-masked Treatment Period and Follow-Up Period and study eye responder definition (study eye responder and fellow eye responder) for the ITT Population. Similar summaries will also be provided separately by tobacco use status (non-user, user), region (US, EU), site, age group (<65 years, ≥65 years), gender, race (white, black or African American, Asian, other), and ethnicity (Hispanic or Latino, not Hispanic or Latino).

In addition, a separate analysis will be provided for time to proptosis response. Time to proptosis response will be measured as the days from the first dose date to the earliest visit date in which the criteria are met for proptosis response (≥ 2 mm reduction from Baseline in proptosis in the study eye, without deterioration [≥ 2 mm increase] of proptosis in the fellow eye) within the Double-Masked Treatment Period, calculated as:

first visit date in which criteria for proptosis response are met – first dose date + 1

Subjects not achieving a proptosis response within the Double-Masked Treatment Period will be censored at their last visit date with a proptosis assessment within the Double-Masked Treatment Period. Kaplan-Meier estimates will be provided by treatment group

of the 25th percentile, median, and 75th percentile. 95% CIs will be provided for the median. Further, a log-rank test stratified by tobacco use status (non-user, user) will be provided. Analyses will be based on the ITT Population. Any ITT subject who is randomized but not dosed will be censored at Day 0.

The analysis will be repeated for the fellow eye (using a definition analogous to the one used for the study eye).

Plots of the time to proptosis response will be provided by treatment group for the study eye and fellow eye.

8.2. SECONDARY EFFICACY ENDPOINTS AND ANALYSES

Secondary endpoints (with the main analyses defined in the study eye as appropriate) will be analyzed using the ITT Population in a hierarchical manner in the order as presented in the following sub-sections. For each outcome measure, teprotumumab will be tested against placebo at the 0.05 significance level only if the test statistic was statistically significantly in favor of teprotumumab for the outcome measure preceding it in the hierarchical order.

8.2.1. CAS and Proptosis Categorical Response

The CAS will be completed at Screening, Day 1 and Weeks 6, 12, 18, and 24 (or PW) during the Treatment Period, and at Weeks 28, 36, 48, 60, and 72 (or PW) during the Follow-Up Period using the 7 item European Group on Graves' Ophthalmopathy (EUGOGO) amended CAS (Mourits et al, 1989) for each eye.

To promote consistency in data collection across clinical trial sites, all investigators will be provided with training and copies of the Clinical Manifestations chapter in Graves' Orbitopathy: *A Multidisciplinary Approach* – Questions and Answers (Dickinson, 2010). Except when strictly unavoidable, the same observer should conduct each CAS evaluation for the full duration of the study.

For each item present, one point is given. The sum of these points is the total score.

- Spontaneous orbital pain
- Gaze evoked orbital pain
- Eyelid swelling that is considered to be due to active (inflammatory phase) TED/GO
- Eyelid erythema
- Conjunctival redness that is considered to be due to active (inflammatory phase) TED/GO (ignore "equivocal" redness)

- Chemosis
- Inflammation of caruncle or plica

The first secondary efficacy endpoint to be tested is the overall responder rate, measured as the percentage of subjects with ≥ 2-point reduction in CAS AND ≥ 2 mm reduction in proptosis from Baseline, provided there is no corresponding deterioration (≥ 2 point/mm increase) in CAS or proptosis in the fellow eye at Week 24.

8.2.1.1. Primary Analysis of CAS and Proptosis Categorical Response

Similar to the primary endpoint analysis, this endpoint will be analyzed using an analysis of the stratified difference in proportions of responders with stratification by tobacco use (non-user, user).

Subjects missing the Week 24 evaluation will be considered to be treatment failures (non-responders) for the analysis.

The null hypothesis tested for this secondary efficacy endpoint is that teprotumumab and placebo achieve equal proportions of responders at Week 24. The corresponding alternative hypothesis is that teprotumumab achieves an unequal proportion of responders at Week 24 compared to placebo. This secondary efficacy endpoint will be achieved if the null hypothesis is rejected at the 0.05 alpha level in favor of teprotumumab and the primary efficacy endpoint was achieved.

Risk differences between teprotumumab and placebo will be estimated along with the corresponding 95% CIs and p-values.

A bar graph will be provided of the proportion of responders for the study eye by visit and treatment group for the ITT Population.

8.2.1.2. Additional Analysis of CAS and Proptosis Categorical Response

The analysis will be repeated using logistic regression with tobacco use as a covariate.

Further, the definition of responder will be applied for each study visit (Weeks 6, 12, 18, and 24), and the analysis of risk difference will be repeated for the ITT Population, considering subjects 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, a calculation of fellow eye response will be conducted (using a definition analogous to the one used for the study eye) at Week 24 and all other study visits (Weeks 6, 12, and 18). The primary

analysis of risk difference will be conducted at Week 24 and all other study (Weeks 6, 12, and 18) visits for the ITT Population, considering subjects missing the evaluations as treatment failures (non-responders) and separately considering the observed results only.

Descriptive summaries of observed responder results for both study eye responder definitions (study eye responder and fellow eye responder) will be provided by treatment group for the ITT Population for each visit in the double-masked Treatment Period and Follow-Up Period. Similar summaries will also be provided separately by tobacco use status (non-user, user), region (US, EU), and site.

8.2.2. CAS Categorical Response

8.2.2.1. Primary Analysis of CAS Categorical Response

The second secondary efficacy endpoint to be tested is CAS as a categorical response variable defined as a reduction to a CAS of 0 or 1 at Week 24. Similar to the primary endpoint analysis, this endpoint will be analyzed using an analysis of the stratified difference in proportions of responders with stratification by tobacco use (non-user, user).

Subjects missing the Week 24 evaluation will be considered to be treatment failures (non-responders) for the analysis.

The null hypothesis tested for this secondary efficacy endpoint is that teprotumumab and placebo achieve equal proportions of responders at Week 24. The corresponding alternative hypothesis is that teprotumumab achieves an unequal proportion of responders at Week 24 compared to placebo.

The risk difference between teprotumumab and placebo will be estimated along with the corresponding 95% CIs and p-values.

A bar graph will be provided of the proportion of responders for the study eye by visit and treatment group for the ITT Population.

8.2.2.2. Additional Analysis of CAS Categorical Response

Analysis of CAS categorical response will be repeated using logistic regression, with tobacco use status as a covariate.

Further, the definition of responder will be applied for each study visit (Weeks 6, 12, 18, 24) and the primary analysis of risk difference will be repeated for the ITT Population, considering subjects 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, a calculation of fellow eye response will be conducted at Week 24 and all other study visits (Weeks 6, 12, and 18) in the double-masked Treatment Period. The primary analysis of risk difference will be conducted at Week 24 and all other study visits (Weeks 6, 12, and 18) in the double-masked Treatment Period for the ITT Population, considering subjects missing the evaluations as treatment failures (non-responders) and separately considering the observed results only.

Descriptive summaries of observed responder results will be provided for the study eye and fellow eye by treatment group for the ITT Population for each visit in the double-masked Treatment Period and Follow-Up Period. Similar summaries will also be provided separately by tobacco use status (non-user, user), region (US, EU), and site.

8.2.3. Proptosis as a Continuous Variable

8.2.3.1. Primary Analysis of Proptosis as a Continuous Variable

For the third secondary efficacy endpoint analysis of proptosis as a continuous variable, an MMRM analysis of covariance (ANCOVA) model will be fit to the individual change from Baseline scores for the study eye, with terms in the model being the Baseline score, tobacco use status (non-user, user), treatment group, visit, and the visit-by-treatment and visit-by-baseline-score interactions. The unstructured (UN) variance-covariance matrix will be used. In the event that this matrix does not allow for model convergence, 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 subjects in the ITT Population 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 subjects 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% CIs and p-value. The main focus of the analysis will be to test the overall treatment difference through Week 24, with the main results consisting of the overall estimated LS means, SEs and their difference with the SE, 95% CI and p-value. The p-value for the overall treatment difference between teprotumumab and placebo will be used in the hierarchical testing of secondary endpoints.

Example SAS code is provided in the Appendix in Section 21.3.

8.2.3.2. Additional Analysis of Proptosis as a Continuous Variable

Estimated LS means, SEs and their differences with 95% CIs and p-values from the MMRM ANCOVA model will be provided for all post-baseline visits. A line graph will be provided plotting the LS mean change from baseline by visit and treatment group for the ITT Population, along with the 95% CIs of the LS means from the model.

Further, the analysis will be repeated using the fellow eye results on the ITT Population.

Descriptive summaries of observed proptosis as a continuous variable will be provided for the study eye and fellow eye by treatment group for the ITT Population for each visit in the double-masked Treatment Period and Follow-Up Period. Similar summaries will be provided separately by tobacco use status (non-user, user), region (US, EU), and site.

8.2.4. Diplopia Categorical Response

At each visit, subjects are assessed for diplopia. The subjective diplopia score (0=no diplopia; 1=intermittent, i.e. diplopia in primary position of gaze, when tired or when first awakening; 2=inconstant, i.e. diplopia at extremes of gaze; 3=constant, i.e. continuous diplopia in primary or reading position) is recorded for each eye. A subject will be considered to have diplopia if a score > 0 is observed in the study eye at baseline.

8.2.4.1. Primary Analysis of Diplopia Categorical Response

The fourth secondary endpoint is diplopia responder. Only subjects with diplopia in the study eye at baseline (diplopia > 0) will be included in this analysis. A subject will be considered a responder if the grade of diplopia (scored from 0 to 3) decreases by at least one grade in the study eye without worsening by at least one grade in the non-study eye. Subjects with diplopia at baseline who are missing the Week 24 diplopia evaluation will be considered treatment failures (non-responders) for the analysis. Similar to the primary endpoint analysis, this endpoint will be analyzed by assessing the stratified difference in proportions of responders with stratification by tobacco use (non-user, user).

The null hypothesis tested for this secondary efficacy endpoint is that teprotumumab and placebo achieve equal proportions of responders for diplopia at Week 24. The corresponding alternative hypothesis is that teprotumumab achieves an unequal proportion at Week 24 compared to placebo.

The risk difference comparing teprotumumab and placebo will be estimated along with the corresponding 95% CIs and p-values.

A bar graph will be provided of the proportion of subjects with improvement in each eye by visit and treatment group for the ITT Population.

8.2.4.2. Additional Analyses of Diplopia Categorical Response

The primary analysis will be repeated using logistic regression with tobacco use status as a covariate.

Further, the definition of responder will be applied for each study visit (Weeks 6, 12, 18, 24) and the analysis of risk difference will be repeated for the ITT Population, considering subjects 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, a calculation of fellow eye response will be conducted at Week 24 and all other study visits (Weeks 6, 12, and 18) in the double-masked Treatment Period. The primary analysis of risk difference will be conducted at Week 24 and all other study visits (Weeks 6, 12, and 18) in the double-masked Treatment Period for the ITT Population, considering subjects missing the evaluations as treatment failures (non-responders) and separately considering the observed results only.

Descriptive summaries of observed responder results will be provided by treatment group for the ITT Population for each visit in the double-masked Treatment Period and Follow-Up Period. Similar summaries will also be provided separately by tobacco use status (non-user, user), region (US, EU), and site.

8.2.5. GO-QoL Questionnaire Overall Score

For the fifth secondary efficacy endpoint analysis, the GO-QoL questionnaire (Terwee et al, 1998) will be completed at Baseline, Weeks 6, 12, 24 (or PW) during the Treatment Period, and at Weeks 28, 48, and 72 (or PW) during the Follow-Up Period. The GO-QoL is a 16-item self-administered questionnaire divided into two subsets and used to assess the perceived effects of TED by the subjects on (i) their daily physical activity as it relates to visual function, and (ii) psychosocial functioning.

The sum of the scores from each set of eight questions will be calculated and transformed to a scale from 0 to 100 – one for visual function (VF), one for appearance (A) and one for the overall combined score. Scores will be transformed as follows:

Transformed score = [(sum of each score – number of completed items) / (2 * number of completed items)] * 100.

8.2.5.1. Primary Analysis of GO-QoL Questionnaire Overall Score

For the overall transformed score an MMRM ANCOVA model will be fit to the individual change from Baseline scores, with terms in the model being the Baseline score, tobacco use status (non-user, user), treatment group, visit, and the visit-by-treatment and visit-by-baseline-score interactions. The unstructured (UN) variance-covariance matrix will be used. In the event that this matrix does not allow for model convergence, the following three variance-covariance matrices will be attempted in order until one converges: TOEPH, CSH, and TOEP. If there are any subjects in the ITT Population 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 subjects 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% CIs and p-value. The main focus of the analysis will be to test the overall treatment difference through Week 24, with the main results consisting of the overall estimated LS means, SEs and their difference with the SE, 95% CI and p-value. The p-value for the overall treatment difference between teprotumumab and placebo will be used in the hierarchical testing of secondary endpoints.

8.2.5.2. Additional Analysis of GO-QoL Questionnaire Overall Score

Estimated LS means, SEs and their differences with 95% CIs and p-values from the MMRM ANCOVA model will be provided for all post-baseline visits. A line graph will be provided plotting the LS mean change from baseline by visit and treatment group for the ITT Population, along with the 95% CIs of the LS means from the model.

Descriptive summaries of the transformed scores will be provided by treatment group for the ITT Population at each visit in the double-masked Treatment Period and Follow-Up Period. Similar summaries will be provided separately by tobacco use status (non-user, user), region (US, EU), and site.

8.3. EXPLORATORY EFFICACY ENDPOINTS AND ANALYSIS

8.3.1. Clinical Measures of Severity

The following items, based on the EUGOGO Consensus Statement (Bartalena et al, 2008; Wiersinga et al, 2006) will be assessed at screening, Day 1, and Weeks 6, 12, 18, 24 (or PW) during the Treatment Period and Weeks 28, 36, 48, 60, and 72 (or PW) during the Follow-Up Period. Except when strictly unavoidable, the same observer should conduct each evaluation of severity measure for the full duration of the study.

8.3.1.1. Categorical Response Variables

Table 2 provides a list of each Clinical Measures of Severity item and the assessment scale for classifying overall response.

Table 2: Clinical Measures of Severity

Clinical Measures of Severity Item and Assessment Scale for Study Eye	Minimum change required for classifying overall response
Exophthalmos (measured in mm using the same Hertel exophthalmometer and same intercanthal distance for each individual subject)	Decrease ≥ 2 mm
Lid aperture (distance between the lid margins in mm with the subject looking in the primary position, sitting relaxed and with distant fixation)	Decrease ≥ 2 mm
Swelling of the eyelids (absent, mild, moderate, severe)	Decrease ≥ One grade
Redness of the eyelids (absent, present)	Decrease ≥ One grade
Redness of the conjunctiva (absent, present)	Decrease ≥ One grade
Conjunctival edema (absent, present)	Decrease ≥ One grade
Inflammation of the caruncle or plica (absent, present)	Decrease ≥ One grade
Subjective diplopia score (0=no diplopia; 1=intermittent, i.e. diplopia in primary position of gaze, when tired or when first awakening; 2=inconstant, i.e. diplopia at extremes of gaze; 3=constant, i.e. continuous diplopia in primary or reading position)	Baseline diplopia score > 0 and decrease ≥ One grade
Eye muscle involvement (ductions in degrees)	Increase ≥ 8° in at least one direction of gaze
Corneal involvement (absent/punctate keratopathy/ulcer)	Decrease ≥ One grade

Clinical Measures of Severity Item and Assessment Scale for Study Eye	Minimum change required for classifying overall response
Optic nerve involvement (best corrected visual acuity, color vision, optic disc, relative afferent pupillary defect (absent, present), and visual fields if optic nerve compression is suspected.	Change of best corrected visual acuity by ≥ 2 lines on Snellen chart, or substantial color vision change, or significant change of visual fields, or significant change in optic disc appearance, or (Dis-) appearance of relative afferent pupillary defect

Response is assessed individually for each item.

The percentages of study eye responders for each item will be analyzed using the same analysis of risk difference method as the primary endpoint based on the ITT Population as described in Section 8.1.1. Subjects missing the evaluation will be considered treatment failures (non-responders) for the analysis. This analysis will be repeated for the fellow eye. Subjects who have a baseline assessment such that improvement is not possible (i.e. they have the best possible assessment) will be considered a responder if they maintain the baseline assessment at a visit for all categorical items except subjective diplopia. The analysis of optic nerve involvement will be generated only if at least 10% of the subjects have optic nerve decompression.

The Clinical Measures of Severity observed results will be summarized for each visit in the double-masked Treatment Period and Follow-Up Period for the study eye and fellow eye by treatment group for the ITT Population with the number and percentage of subjects being classified as responders on each individual item.

8.3.1.2. Motility Restriction

Motility is examiner assessed by estimating the degrees of restriction in eye movements. Monocular excursions in horizontal and vertical directions of gaze are recorded using the light reflex (LR) (Dolman et al, 2012) test.

The clinician will shine a pen light in line with the eye being examined in ambient room light and observe the subject's eye along the light's axis. The subject will be asked to look in the four cardinal directions and the position of the light reflex is viewed on the surface of the cornea. If the light touches the limbus, the eye is assessed to be turned 45 degrees, if half way between the limbus and pupil edge, the eye is assessed at 30 degrees, and if it is at the pupil edge, it was assessed at 15 degrees. Intermediate ductions are judged by estimating the light position between these points to the nearest 5 degrees.

The monocular ductions of each eye (degrees) will be recorded for adduction, abduction, supraduction and infraduction. For each motility restriction measure, the study eye data will be analyzed using the same MMRM ANCOVA model as the secondary

endpoint of proptosis as a continuous variable as described in Section 8.2.3 based on the ITT Population. This analysis will be repeated for the fellow eye.

Descriptive summaries for the study eye and fellow eye will be provided for each motility measure by treatment group for the ITT Population for each visit in the double-masked Treatment Period and Follow-Up Period.

8.3.2. CAS as a Continuous Variable

For the analysis of CAS as a continuous variable, the study eye data will be analyzed using the same MMRM ANCOVA model as the secondary endpoint of proptosis as a continuous variable as described in Section 8.2.3 based on the ITT Population. This analysis will be repeated for the fellow eye.

Descriptive summaries of CAS as a continuous variable for the study eye and fellow eye will be provided by treatment group for the ITT Population for each visit in the double-masked Treatment Period and Follow-Up Period.

8.3.3. Categorical Overall Response

To further explore the response based on both proptosis and CAS reduction, each subject will be classified into one of four response categories at Week 24:

- High responders: Subjects who had a reduction in both proptosis and CAS of 3 or more (≥ 3) from Baseline in the study eye, and no deterioration in the fellow eye (i.e., increase in CAS ≥ 2 points or increase in proptosis ≥ 2 mm).
- Responders: Subjects who had a reduction in both CAS and proptosis of 2 or more (but less than 3) from Baseline in the study eye, and no deterioration in the fellow eye.
- Low Responders: Subjects who had a reduction in both CAS and proptosis of 1 or more (but less than 2) from Baseline in the study eye, and no deterioration in the fellow eye.
- Non-Responders: Subjects who did not fit into any of the above categories, or were not present for the Week 24 evaluation

A proportional odds logistic model for ordered categorical data with treatment group as the model effect and tobacco use status (non-user, user) as a stratification factor will be used to analyze the data for the study eye. The estimate of the common odds ratio comparing teprotumumab to the placebo will be provided along with the corresponding 95% CI and p-value. In addition, the p-value testing the appropriateness of the proportional odds assumption will be provided.

Example SAS code is provided in the Appendix in Section 21.4.

The analysis will be repeated for the fellow eye, using analogous definitions of response categories.

Descriptive summaries will be provided for both response definitions (study eye and fellow eye response) by treatment group for the count and percentage of subjects in each category for the ITT Population.

8.3.4. GO-QoL Questionnaire (Visual Function and Appearance Scale Scores)

The MMRM ANCOVA analysis conducted for the GO-QoL Questionnaire overall score as described in Section 8.2.4, will be repeated for the transformed VF scale score and A scale score separately based on the ITT Population.

Descriptive summaries of the transformed VF scale scores and transformed A scale scores will be provided by treatment group for the ITT Population at each visit in the double-masked Treatment Period and Follow-Up Period.

8.3.5. Percent Change from Baseline in Proptosis

For the analysis of percent change from Baseline in proptosis, the study eye data will be analyzed using the same MMRM ANCOVA model as the secondary endpoint of proptosis as a continuous variable as described in Section 8.2.3 based on the ITT Population. This analysis will be repeated for the fellow eye.

Descriptive summaries of percent change from baseline in proptosis for the study eye and fellow eye will be provided by treatment group for the ITT Population for each visit in the double-masked Treatment Period and Follow-Up Period.

8.3.6. Time to Relapse in Follow-Up Period

Time to relapse in the Follow-Up Period (increase in proptosis >=2 in the study eye since Week 24 or increase in CAS ≥ 2 points since Week 24 with an absolute CAS >=4 in the study eye following Week 24 visit) will be measured as the days from the Week 24 visit date to the date in each relapse criteria are met, calculated as:

visit date in which relapse criteria are met – Week 24 visit date + 1

The analysis will be based on subjects who enter the Follow-Up Period (i.e. complete or discontinue study after Week 24). Subjects not experiencing a relapse will be censored at their last visit date with a proptosis assessment in the Follow-Up Period. Kaplan-Meier estimates will be provided by treatment group of the 25th percentile, median, and 75th percentile.

9. ANALYSIS OF PHARMACOKINETICS

PK blood samples will be collected at pre-dose (within 1 hour prior to infusion) and the end of infusion at Day 1, Weeks 3 and 9, as well as single samples at Weeks 1, 4, and 24.

Length of infusion will be approximately 90 min for infusion 1 (Day 1) and 2 (Week 3) while 60 min for remaining doses. Pre-dose PK sampling will be considered as time 0 and the end of infusion time will be recorded.

The following presentations of subject serum concentration data covered in this SAP will be provided for teprotumumab for the PK Population:

- A listing including subject, week/time point (actual, planned), treatment and serum concentrations. Actual end of infusion sampling times are expressed relative to the start time of infusion.

- A table summary of serum concentrations at each time point (n; mean, SD, coefficient of variation (CV)% calculated as $100\% \times SD/mean$, minimum, 25^{th} percentile, median, 75^{th} percentile and maximum)

Concentrations below the limit of quantification (BLQ) collected on Day 1 pre-dose will be summarized as zero. All other concentrations BLQ will be excluded from the analysis summaries.

An additional analysis of PK is described in Section 11.4.

10. ANALYSIS OF BIOMARKERS

Serum samples will be obtained prior to dosing at Day 1, Weeks 12 and 24 and may be analyzed for biomarkers including, but not limited to IL-4, IL-6, IL-10, IL-12, IL-13, IL-17, IL-23, IL-1β, sIL-1RA, INFγ, TGFβ, TNFα, micRNA and possibly functional thyroid stimulating hormone receptor stimulating, blocking, and binding antibody (TSH-R-Ab). The data will be presented as follows:

- A listing including the actual sampling time of biomarker blood sample collection for each subject, including the deviation in time from the protocol scheduled time, if applicable
- A listing including serum biomarkers concentration data and change from Baseline by subject, time point, and treatment (in relevant concentration units)
- A table summary of biomarker data and change from Baseline at each time point by treatment group (n; mean, SD, CV% calculated as 100% × SD/mean, minimum, 25th percentile, median, 75th percentile, and maximum)
- Mean ± SD chronological profiles (linear scale) for biomarkers (absolute concentration values and fold change from Baseline) by treatment group.

Missing concentration values for subjects who were administered the scheduled study treatments will be considered as non-informative missing. No concentration estimates will be provided for missing sample values. Mean serum concentrations will not be presented if 50% or more of the actual values at any one time point are BLQ or missing within the treatment group.

11. SAFETY

Safety analyses will be based on the Safety Population. Safety will be assessed on the basis of AEs, concomitant medication use (refer to Section 7.4), immunogenicity testing, physical examinations, ophthalmic examinations, vital signs, clinical safety laboratory evaluations, pregnancy tests, and ECGs. All safety information will be provided in subject listings.

11.1. EXTENT OF EXPOSURE

Study drug exposure will be measured as the number of study drug doses administered (including any incomplete doses), the number of days on drug, and the number of days on study (double-masked Treatment Period, Follow-up Period, and overall). The number of doses administered will be summarized by treatment group as the count and percentage of subjects receiving each number of doses, as well as a continuous variable. Further, the number of days on drug and on study will be summarized descriptively by treatment group and calculated as follows:

- Number of days on drug = last dose date – first dose date + 1
- Number of days on study (double-masked Treatment Period):
    - For subjects who discontinue during the double-masked Treatment Period = discontinuation date – first dose date + 1
    - For subjects who complete the double-masked Treatment Period = Week 24 visit date – first dose date + 1
- Number of days on study (Follow-Up Period):
    - Subjects who do not enter the Follow-Up Period (i.e., complete or discontinue study on or prior to Week 24) will not contribute to this analysis
    - For subjects who participate in the Follow-Up Period = completion/discontinuation date – Week 24 visit date + 1
- Number of days on study (overall) = completion/discontinuation date – first dose date + 1

11.2. TREATMENT COMPLIANCE

A descriptive summary will be provided for the count and percentages of subjects with each of the following at any infusion visit:

- Planned doses that were not administered completely

- Infusion interruptions

Summaries will be provided by treatment group.

11.3. ADVERSE EVENTS

All adverse events will be coded using MedDRA version 20.1. The TEAE reporting period begins with administration of the first dose of study medication on Day 1 and continues until 3 weeks (21 days) after the last dose of study drug. 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 (Week 72 or PW).

Missing data conventions for AEs are described in Section 6.3.4.

An overall summary of TEAEs will be provided by treatment group, including the number and percentage of subjects with each AE type as well as the number of events for each of the following:

- TEAEs
- Serious TEAEs
- Related TEAEs
- Related serious TEAEs
- TEAEs with an intensity of severe or higher
- TEAEs leading to interruption of study drug administration
- TEAEs leading to permanent withdrawal of study drug
- Related TEAEs leading to permanent withdrawal of study drug
- TEAEs leading to study discontinuation
- Related TEAEs leading to study discontinuation
- TEAEs leading to death

Additional TEAE summaries will be provided by treatment group, including the number and percentage of subjects experiencing TEAEs for the following:

- TEAEs overall and by SOC and PT, including the following subgroup analyses:

- Age (<65 years, ≥65 years)
- Gender (male, female)
- Race (white, black or African American, Asian, other)
- Ethnicity (Hispanic or Latino, not Hispanic or Latino)

- Related TEAEs overall and by SOC and PT
- TEAEs by maximum intensity, overall and by SOC and PT
- Serious TEAEs, overall and by SOC and PT
- Related TEAEs by intensity, overall and by SOC and PT
- TEAEs leading to permanent withdrawal of study drug, overall and by SOC and PT
- TEAEs of special interest including potential infusion-related reaction, anaphylactic reaction, hearing loss, hyperglycemia, muscle spasm, and diarrhea (see Section 11.3.1), by SOC and PT
- TEAEs of special interest with missing AE start times including potential infusion-related reaction of any type, infusion-related reaction excluding those of an anaphylactic nature, and anaphylactic reaction with missing AE start times (see Section 11.3.1), by SOC and PT

For summaries by SOC, PT, and maximum intensity, a subject will only be counted once for each SOC based on the maximum intensity level reported for that SOC and once for each unique PT within that SOC level at the maximum intensity level reported for that PT. For summaries by SOC and PT only, a subject will be counted at most once at the SOC level and at most once at each unique PT within the SOC level. Summaries presenting the frequency of TEAEs by SOC and PT will be ordered alphabetically by SOC and then, within a SOC, alphabetically by PT. Any AEs that accumulate after the unmasking of the treatment groups at the end of the double-masked Treatment Period will be subsequently identified prior to generation of the final analysis.

The aforementioned summaries will be repeated for AEs occurring during the Follow-Up Period with the exception of AEs leading to permanent withdrawal of study drug and the subgroup analyses, which will be provided for the TEAEs only. Denominators for the Follow-Up Period AE summaries will be based on the number of subjects entering the Follow-Up Period (i.e., complete or discontinue study after Week 24).

In addition to the listing of all AEs, separate listings will be provided of serious AEs, AEs leading to withdrawal of study drug, AEs leading to study discontinuation, AEs of special interest, and AEs leading to death. TEAEs will be identified on each listing.

11.3.1. Adverse Events of Special Interest

Adverse events of special interest will be identified as follows:

- Potential infusion-related reactions: TEAEs that occur within 2 hours after initiating infusion
- Anaphylactic reactions: TEAEs that occur within 2 hours after initiating infusion AND meet any one of the 3 criteria from Section 2.6.3 of the Introductory Guide for Standardised MedDRA Queries (SMQs) (MedDRA, 2017):
    - A narrow term or a term from Category A
    - A term from Category B AND a term from Category C
    - A term from Category D AND [a term from Category B OR a term from Category C]
- Hyperglycemia: TEAEs by any PTs under the SMQ "Hyperglycaemia/new onset diabetes mellitus" (narrow).
- Muscle spasm: TEAEs by the PT "muscle spasm"
- Diarrhea: TEAEs by any PTs under the SMQ "Noninfectious diarrhea" (broad search)
- Hearing loss: TEAEs by any PTs under the SMQ "Hearing impairment" (a sub-SMQ under "Hearing and vestibular disorders") and HLT "Hearing losses"
- Potential infusion-related reactions of any type that are missing the onset time: TEAEs that are missing the AE start time, occur on the day of an infusion and meet one of the following criteria:
    - SMQ Angioedema (broad)
    - SMQ Hypersensitivity (broad)
    - SMQ Hypertension (narrow)
    - SMQ Anaphylactic reactions - meet any of the 3 criteria from Section 2.6.3 of the Introductory Guide for SMQs (MedDRA, 2017):
        - A narrow term or a term from Category A
        - A term from Category B AND a term from Category C
        - A term from Category D AND [a term from Category B OR a term from Category C]

  - HLTs: “Rashes, eruptions and exanthems NEC”, “Erythemas”, “Pruritus NEC”
  - PTs: Pyrexia, Flushing, Feeling hot, Chest pain, Chest discomfort, Myalgia, Myalgia intercostal, Back pain, Dizziness

- Potential infusion-related reactions excluding those of an anaphylactic nature that are missing the onset time: TEAEs that are missing the AE start time, occur on the day of an infusion and meet one of the following criteria:

  - SMQ Angioedema (broad)
  - SMQ Hypersensitivity (broad)
  - SMQ Hypertension (narrow)
  - HLTs: “Rashes, eruptions and exanthems NEC”, “Erythemas”, “Pruritus NEC”
  - PTs: Pyrexia, Flushing, Feeling hot, Chest pain, Chest discomfort, Myalgia, Myalgia intercostal, Back pain, Dizziness

- Anaphylactic reactions that are missing the onset time: TEAEs that are missing the AE start time, occur on the day of an infusion AND meet any one of the 3 criteria from Section 2.6.3 of the Introductory Guide for Standardised MedDRA Queries (SMQs) (MedDRA, 2017):

  - A narrow term or a term from Category A
  - A term from Category B AND a term from Category C
  - A term from Category D AND [a term from Category B OR a term from Category C]

11.4. IMMUNOGENICITY

Serum samples will be collected prior to dose at Day 1, Weeks 3, 9, 24, 36 and 72 (or PW). If a subject tests positive for ADA after confirmatory and reactive titer testing, the sample will then be tested for neutralizing antibody (NAb). If the subject tests positive for NAb, he/she will be followed until levels either revert to Baseline or the subject’s value decreases or remains stable.

Immunogenicity endpoints include:

- Incidence of ADA and NAb (out of positive ADA samples) by visit and treatment group

Overall positive ADA result for a subject will be defined as at least one positive antibody. Cumulative negative result will be defined as negative ADA result at all time points for a subject.

Immunogenicity sampling times and results will be listed.

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

11.5. PHYSICAL EXAMINATION

A physical examination will be performed at Screening, Baseline (Day 1) and thereafter at Weeks 1, 6, 12, 18, 24 (or PW) during the Treatment Period and at Weeks 48 and 72 (or PW) of the Follow-Up Period. The physical examination will include assessment of presence or absence of pretibial myxedema on Day 1, Week 24 (or PW) of the Treatment Period, and Week 72 (or PW) of the Follow-Up Period. If present, measurements of instep and calf will be taken. Height will be measured at Screening only.

A shift table will be presented providing the count of subjects with a presence or absence of pretibial myxedema at Baseline compared to each post-baseline visit for each side by treatment group with percentages based on subjects with a non-missing value at the Baseline and post-baseline visit. Further, an overall shift table will be provided capturing the worst assessment between the right and left side at Baseline compared to each post-baseline visit.

11.6. OPHTHALMIC EXAMINATION

A complete undilated ophthalmic examination will be performed as part of the physical examination. The ophthalmic exam should include best-corrected visual acuity, pupil exam, color vision assessment, Ishihara color plates (or equivalent) or related red desaturation, intraocular pressure and slit lamp exam. If significant abnormalities are noted compared with previous visits, including a loss of 2 lines or more of vision, development of pupil abnormalities including APD, rise in intraocular pressure, development of corneal infiltrates or other abnormalities not here specified but of concern to the ophthalmologist, further investigations of visual function will be conducted according to the ophthalmologist decision.

New findings reported from on-study ophthalmic examinations will not be reported as AEs if according to the investigator the abnormalities are related to TED and not related to the study drug.

For each of the assessments (pupil exam, color vision assessment, intraocular pressure, and slit lamp exam), shift tables will be presented providing the count of subjects with

each type of finding (normal, abnormal – not clinically significant [NCS], or abnormal – clinically significant [CS]) at Baseline compared to each post-baseline visit by treatment group with percentages based on subjects with a non-missing value at the Baseline and post-baseline visit. In addition, a summary table will be provided for loss of 2 lines or more of vision (yes or no) at each post-baseline visit.

11.7. VITAL SIGNS

Blood pressure and pulse measurements will be obtained with the subject's arm unconstrained by clothing or other material and while the subject is sitting up. When possible, the same arm will be used for measurements in all study visits. Temperature will be obtained. Assessment of vital signs, including weight, heart rate, systolic blood pressure, diastolic blood pressure, respiratory rate, and temperature will be performed at all clinic visits. On Day 1 and Week 3, vital signs will be measured at pre- and post-infusion. For all other visits, vital signs will be measured prior to infusion only. In addition, if immediate infusion-associated events are noted during the infusion, vital signs will be monitored every 5 minutes until stable and then every 15 minutes for 2 additional determinations. Also, vital signs will be monitored every 15 minutes from the start of the infusion through 60 minutes after infusion completion for any subsequent infusions after the previous occurrence of immediate or delayed infusion-associated events.

Descriptive summaries of observed and change from Baseline values will be presented for each vital sign parameter by treatment group and visit. Vital sign measurements that are monitored as a result of an infusion-associated event as described above will not be included in the descriptive summaries but will be presented in subject listings.

The following conversion factor will be used to convert any temperatures reported in degrees Fahrenheit to Celsius:

Temperature (in °C) = 5/9 (Temperature [in °F]-32).

11.8. CLINICAL LABORATORY SAFETY EVALUATIONS

With the exception of urine pregnancy tests, a central study laboratory will be used for all protocol-specified clinical laboratory parameters. Table 3 provides the schedule of collection.

Table 3 Schedule of Clinical Laboratory Safety Tests, Including Thyroid Panel and Hyperglycemia Monitoring

Analysis Panel		Treatment Period											Follow-Up Period				
	SCR	BL	W1	W3	W4 M1	W6	W9	W12 M3	W15	W18	W21	W24 M6	W28 M7	W36 M9	W48 M12	W60 M15	W72 M18
Chemistry (excl. glucose)	X [1]	X		X		X	X	X		X		X		X			X
Thyroid (FT3, FT4, TSH) [2]	X	X		X		X	X	X		X		X		X			X
Hematology	X	X	X	X	X	X	X	X	X	X	X	X		X			X
Glucose [3]	X	X	X	X	X	X	X	X	X	X	X	X		X			X
HbA1C [4]	X							X				X		X			X
Urinalysis	X	X		X		X	X	X		X		X		X			X
Serum pregnancy [5]	X																
Urine pregnancy [5]		X		X		X	X	X	X	X	X	X	X	X	X		X [6]

BL=Baseline; FT3=free triiodothyronine; FT4=free thyroxine; HbA1C=glycated hemoglobin A1C; M=Month; SCR=Screening; TSH=thyroid stimulating hormone; W=Week.

1. ALT/AST must be ≤ 3 x the ULN and serum creatinine must be < 1.5 x the ULN according to age for randomization.
2. Subjects must be euthyroid with the baseline disease under control or have mild hypo- or hyperthyroidism (defined as FT4 and FT3 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 clinical trial.
3. Non-diabetic subjects will fast at Weeks 1 and 4 only. Diabetic subjects will fast for each blood glucose evaluation. NOTE: Subjects with severe hyperglycemia that does not abate to mild or moderate intensity with anti-diabetic treatment (dose may be skipped up to 2 times prior to permanently discontinuing study drug) will be permanently discontinued from study drug.
4. HbA1C must be < 9.0% for randomization. If the HbA1C is elevated and considered clinically significant at any time point after Screening, it will be repeated approximately every 45 days until it returns to normal or baseline value.
5. Perform for female subjects 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]).
6. Perform for female subjects of childbearing potential who enter the Follow-Up Period but discontinue study participation prior to Week 48.

Laboratory results collected in conventional units will be converted to International System of Units (SI) for all summaries and listings. Clinical laboratory test results (hematology, chemistry, thyroid panels, and urinalysis) and their changes from Baseline will be summarized by visit and treatment group using descriptive statistics. If a continuous laboratory value is reported as either below or above the limits of quantification, the qualifiers will be dropped and the numeric value used in the analysis (e.g., "< 3" will be summarized as "3" and "> 200" will be summarized as "200"). For hematology, chemistry and thyroid panel, results will be categorized as low, normal, or high based on their normal ranges. For urinalysis tests, results will be classified as normal or abnormal. Results out of range will be identified as such on subject listings. Shift tables using categories of low, normal, and high, comparing laboratory test results from Baseline to each visit will be presented with percentages based on subjects with a

non-missing value at Baseline and post-baseline visit. Additionally, a shift table for glucose by Common Terminology Criteria for Adverse Events (CTCAE) grade and visit will be presented. Summaries will be provided separately for hyperglycemia and hypoglycemia.

Box plots will be provided for the observed FT3, FT4, TSH and HbA1C values by treatment group at baseline and Week 24. Separate box plots will be provided for the minimum post-baseline values, minimum change from baseline values, maximum post-baseline values, and maximum change from baseline values for FT3, FT4, TSH and HbA1C values by treatment group.

11.9. PREGNANCY TEST

Pregnancy test results will be provided in a listing only.

11.10. ECG

12-lead ECGs will be performed at Screening, Baseline and Weeks 3, 6, 12, 24 (or PW) of the Treatment Period, and Week 72 (or PW) of the Follow-Up Period. At infusion visits, ECGs will be performed prior to the infusion. The results will be recorded as normal or abnormal on the eCRF and all abnormal results will be evaluated as CS or NCS by the investigator.

Descriptive summaries of observed and change from Baseline values will be presented for each ECG parameter by treatment group and visit, including RR, PR, QRS, QT, QTc, and QTcF.

ECG shift tables will be presented providing the count of subjects with each type of finding (normal, abnormal – NCS, or abnormal – CS) at Baseline compared to each post-baseline visit by treatment group with percentages based on subjects with a non-missing value at the Baseline and post-baseline visit.

Further, a summary will be provided of the count and percent of subjects with any post-Baseline assessment in the following categories:

- QTcF >450 msec (males) or > 470 msec (females)
- QTcF >500 msec

Percentages will be based on the number of subjects with at least one post-Baseline value. Similar summaries will be provided by visit with the denominator based on the number of subjects with data at the given visit.

12. INTERIM ANALYSES

There are no planned interim analyses.

13. CHANGE FROM ANALYSIS PLANNED IN PROTOCOL

Protocol Version 4.0 (31-Jan-2019) specified that the safety analysis population will include all subjects who receive at least one dose and had at least one post-dose safety assessment. This definition has been changed to require only that the subject received at least one dose in order to be considered in the Safety Population.

Further, Protocol Version 4.0 specified that the continuous secondary endpoints (proptosis and GO-QoL) would be analyzed using an MMRM ANCOVA model with the main focus of the analysis being the treatment differences at Week 24, with the main results consisting of the Week 24 estimated LS means and their differences. While those results will still be provided, the main focus of this analysis has been revised to the overall treatment differences through Week 24, with the main results consisting of the overall estimated LS means, SEs and their differences with the SE, 95% CI and p-value.

14. REFERENCE LIST

Bartalena L, Baldeschi L, Dickinson A, et al. Consensus statement of the European Group on Graves' orbitopathy (EUGOGO) on management of GO. European journal of endocrinology. 2008; 158(3): 273-85.

Dickinson AJ. (2010). Clinical manifestations. In W. M. Wiersinga & G. J. Kahaly (Eds.), *Graves' orbitopathy: a multidisciplinary approach - questions and answers* (2nd ed.): Karger Basel.

Dolman PJ, Cahill K, Czyz CN, et al. Reliability of estimating ductions in thyroid eye disease: an International Thyroid Eye Disease Society multicenter study. Ophthalmology. 2012; 119(2): 382-9.

International Council for Harmomisation of Technical Requirements for Pharmaceuticals for Human Use, ICH Harmonised Guideline, "Estimands and Sensitivity Analysis in Clinical Trials," E9(R1), June 16, 2017.

MedDRA: Introductory Guide for Standardised MedDRA Queries (SMQs) Version 20.1 (Sep 2017). Available at: https://www.meddra.org/sites/default/files/guidance/file/smq_intguide_20_1_english.pdf

Mourits MP, Koornneef L, Wiersinga WM, et al. Clinical criteria for the assessment of disease activity in Graves' ophthalmopathy: a novel approach. The British journal of ophthalmology. 1989; 73(8): 639-44.

Terwee CB, Gerding MN, Dekker FW, et al. Development of a disease specific quality of life questionnaire for patients with Graves' ophthalmopathy: the GO-QOL. The British journal of ophthalmology. 1998; 82(7): 773-9.

Wiersinga WM, Perros P, Kahaly GJ, et al. Clinical assessment of patients with Graves' orbitopathy: the European Group on Graves' Orbitopathy recommendations to generalists, specialists and clinical researchers. European journal of endocrinology. 2006; 155(3): 387-9. ''

15. PROGRAMMING CONSIDERATIONS

All tables, data listings, figures (TLFs), and statistical analyses will be generated using SAS® for Windows, Release 9.4 or higher (SAS® Institute Inc., Cary, NC, USA). Computer-generated table, listing and figure output will adhere to the following specifications.

15.1. GENERAL CONSIDERATIONS

- One SAS program can create several outputs.
- Each output will be stored in a separate file.
- Output files will be delivered in RTF format. Two PDF files will be provided for the final tables and listings separately.
- Numbering of TFLs will follow International Conference on Harmonization (ICH) E3 guidance

15.2. TABLE, LISTING, AND FIGURE FORMAT

15.2.1. General

- All TLFs will be produced in landscape format.
- All TLFs will be produced using the Courier New font, size 8
- The data displays for all TLFs will have a minimum 1-inch margin on all 4 sides.
- Headers and footers for figures will be in Courier New font, size 8.
- Legends will be used for all figures with more than 1 variable, group, or item displayed.
- TLFs will be in black and white (no color)
- Specialized text styles, such as bolding, italics, borders, and shading will not be used in the TLFs, unless otherwise specified. On some occasions, superscripts 1, 2, or 3 may be used.
- Only standard keyboard characters will be used in the TLFs. Special characters, such as non-printable control characters, printer-specific, or font-specific characters, will not be used. Hexadecimal-derived characters will be used, where possible, if they are appropriate to help display math symbols (e.g., µ). Certain subscripts and superscripts (e.g., m^2, C_{trough}) will be employed on a case-by-case basis.
- Mixed case will be used for all titles, footnotes, column headers, and programmer-supplied formats, as appropriate.

15.2.2. Headers

- All output should have the following header at the top left of each page:

Horizon Pharma USA, Inc.
Protocol No: HZNP-TEP-301

- All output should have Page n of N at the top or bottom right corner of each page. TLFs should be internally paginated in relation to the total length (i.e., the page number should appear sequentially as page n of N, where N is the total number of pages in the table).
- The date output was generated should appear along with the program name as a footer on each page.

15.2.3. Display Titles

- Each TLF should be identified by the designation and a numeral. A decimal system (x.y and x.y.z) should be used to identify TLFs with related contents. The title is centered. The analysis set should be identified on the line immediately following the title. The title and table designation are single spaced. A solid line spanning the margins will separate the display titles from the column headers. There will be 1 blank line between the last title and the solid line.

Table x.y.z
First Line of Title
Second Line of Title if Needed
ITT Population

15.2.4. Column Headers

- Column headings should be displayed immediately below the solid line described above in initial upper-case characters.
- In the case of efficacy tables, the variable (or characteristic) column will be on the far left followed by the treatment group columns and overall column (if applicable). P-values may be presented under the treatment group columns or in a separate p-value column (if applicable).
- For numeric variables, include “unit” in column or row heading when appropriate.
- Analysis set sizes will be presented for each treatment group in the column heading as (N=xx) (or in the row headings if applicable). This is distinct from the ‘n’ used for the descriptive statistics representing the number of subjects in the analysis set.
- The order of treatments in the tables and listings will be “Placebo” followed by “HZN-001” (teprotumumab 10 mg/kg first infusion and teprotumumab 20 mg/kg remaining 7 infusions).

15.2.5. Body of the Data Display

15.2.5.1. General Conventions

Data in columns of a table or listing should be formatted as follows:

- alphanumeric values are left-justified;
- whole numbers (e.g., counts) are right-justified; and
- numbers containing fractional portions are decimal aligned.

15.2.5.2. Table Conventions

- Units will be included where available
- If the categories of a parameter are ordered, then all categories between the maximum and minimum category should be presented in the table, even if n=0 for all treatment groups in a given category that is between the minimum and maximum level for that parameter. For example, the frequency distribution for symptom severity would appear as:

Severity Rating	n
severe	0
moderate	8
mild	3

- If the categories are not ordered (e.g., Medical History, Reasons for Discontinuation from the Study, etc.), then only those categories for which there is at least one subject represented in one or more groups should be included.
- An unknown or missing category should be added to any parameter for which information is not available for one or more subjects.
- Unless otherwise specified, the estimated mean and median for a set of values should be printed out to one more significant digit than the original values, and standard deviations should be printed out to two more significant digits than the original values. The minimum and maximum should report the same significant digits as the original values. For example, for systolic blood pressure:

n	XX
Mean (SD)	XXX.X (X.XX)
Median	XXX.X
Min, Max	XX, XX

- P-values should be output in the format: “0.xxx”, where xxx is the value rounded to 3 decimal places. Any p-value less than 0.001 will be presented as “<0.001”. Any p-value greater than 0.999 will be presented as “>0.999”.
- Percentage values should be printed to one decimal place, in parentheses with no spaces, one space after the count (e.g., 7 (12.8%), 13 (5.4%)). Unless otherwise

noted, for all percentages, the number of subjects in the analysis set for the treatment group who have an observation will be the denominator. The denominator for percentages will be identified in footnotes. Percentages after zero counts should not be displayed and percentages equating to 100% should be presented as 100%, without any decimal places.

- Missing descriptive statistics or p-values which cannot be estimated should be reported as "NE" for not estimable.
- For categorical summaries (number and percentage of subjects) where a subject can be included in more than one category, footnotes and/or programming notes will identify the selection criteria.
- Where a category with a subheading (such as system organ class) has to be split over more than one page, the subheading will be outputted followed by "(cont)" at the top of each subsequent page. The overall summary statistics for the subheading will only be output on the first relevant page.

15.2.5.3. Listing Conventions

- Listings will be sorted for presentation in order of treatment groups as above, subject number, visit/collection day, and visit/collection time.
- Missing data should be represented on subject listings as either a hyphen ("-") with a corresponding footnote ("- = unknown or not evaluated"), or as "NA", with the footnote "NA = not applicable", whichever is appropriate.
- Dates should be printed in SAS® DATE9.format ("ddMMMyyyy": 01JUL2000). Missing portions of dates should be represented on subject listings as "UN" for missing days and "UNK" for missing months (e.g. UNJUL2000, UNUNK2000). Dates that are missing because they are not applicable for the subject are output as "NA", unless otherwise specified.
- All observed time values must be presented using a 24-hour clock HH:MM or HH:MM:SS format (e.g., 11:26:45, or 11:26). Time will only be reported if it was measured as part of the study.
- Units will be included where available

15.2.5.4. Figure Conventions

- Unless otherwise specified, for all figures, study visits will be displayed on the X-axis and endpoint (e.g., treatment mean change from Baseline) values will be displayed on the Y-axis.

15.2.6. Footnotes

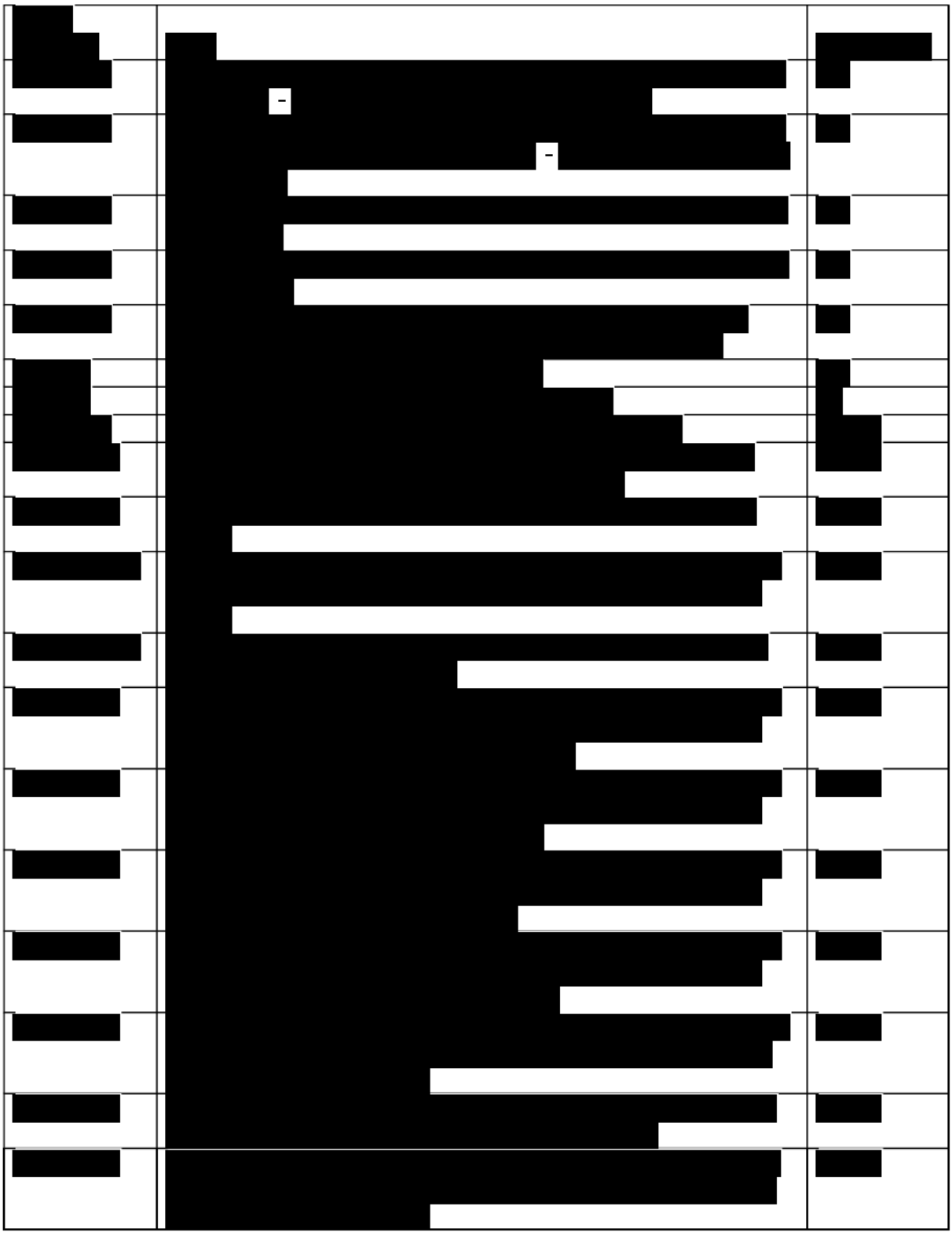
- A solid line spanning the margins will separate the body of the data display from the footnotes.
- All footnotes will be left justified with single-line spacing immediately below the solid line underneath the data display.
- Footnotes should always begin with "Note:" if an informational footnote, or [1], [2], [3], etc. if a reference footnote. Each new footnote should start on a new line where possible.
- Subject specific footnotes should be avoided, where possible.
- Footnotes will be used sparingly and must add value to the table, figure, or data listing. If more than six lines of footnotes are planned, then a cover page may be used to display footnotes, and only those essential to comprehension of the data will be repeated on each page.
- The last line of the footnote section will be a standard source line that indicates the name of the program used to produce the data display, date the program was run, and the TLF source (i.e., 'Program : myprogram.sas Listing source: 16.x.y.z').

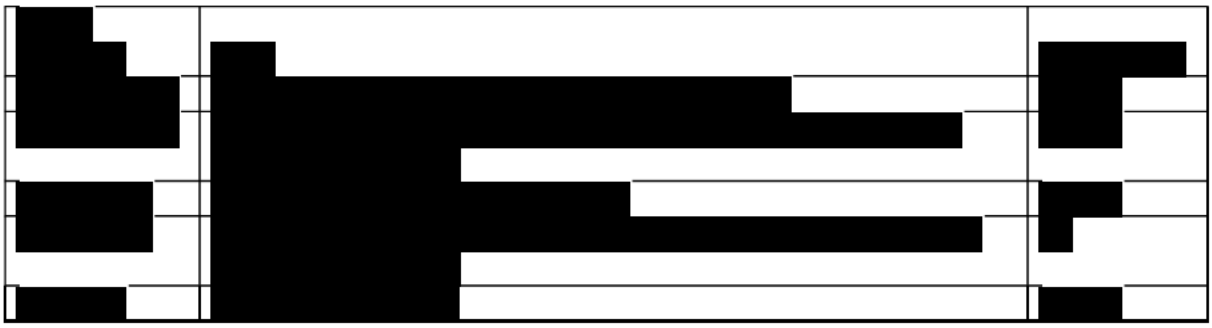
16. QUALITY CONTROL

SAS programs are developed to produce output such as analysis data sets, summary tables, data listings, figures or statistical analyses. INC Research Standard Operating Procedure (SOP) 03.010 and 03.013 provide an overview of the development of such SAS programs.

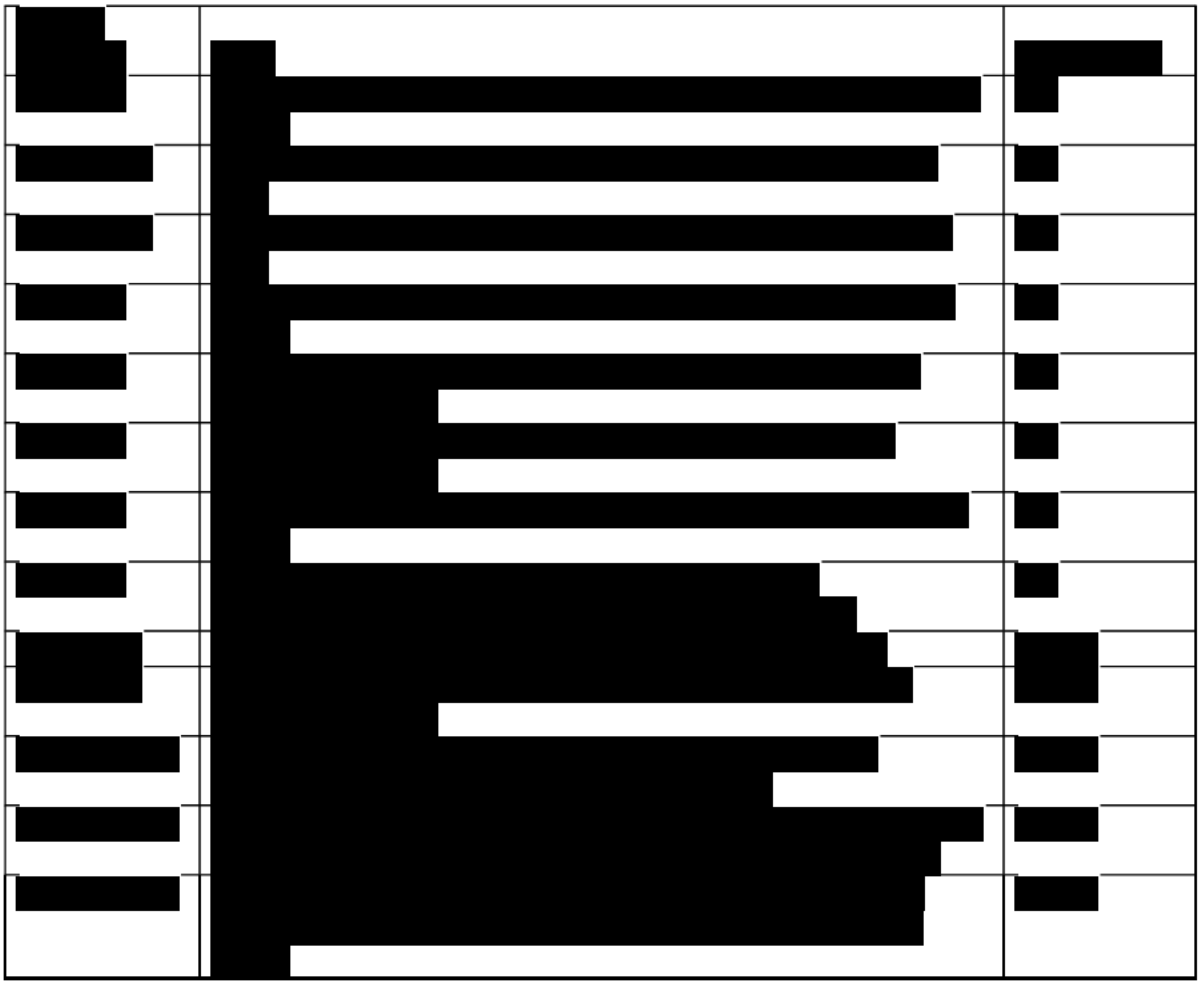
INC Research SOP 03.009 describes the quality control procedures that are performed for all SAS programs and output. Quality control is defined here as the operational techniques and activities undertaken to verify that the SAS programs produce the output by checking for their logic, efficiency and commenting and by review of the produced output.

17. INDEX OF TABLES





18. INDEX OF FIGURES



19. INDEX OF LISTINGS

20. MOCK-UPS

Table, figure and listing mock-ups are provided in a separate file.

21. APPENDIX A – SAS CODE

21.1. PRIMARY ANALYSIS OF PRIMARY ENDPOINT

The following example code will be used to generate the primary analysis results:

```
ods output CommonPdiff=cpdiff;

PROC FREQ DATA = datasetname;
   TABLE TOBACCO * TRTPN * RESPONSE / RISKDIFF(COMMON);
RUN;
```

where RESPONSE = proptosis responder status at Week 24 ('Y' or 'N'); TOBACCO = tobacco use status ('Y' or 'N'); TRTPN = treatment group (1 = Placebo, 2 = Teprotumumab)

The output dataset will be used to calculate the p-value, using code similar to this:

```
data cpdiff; set cpdiff;
   p_value=min(1, 2*(1-probnorm(abs(value/stderr))));
run;
```

21.2. SUPPORTIVE ANALYSIS OF PRIMARY ENDPOINT

A supportive analysis of the primary endpoint will use logistic regression, with code such as the following:

```
PROC GENMOD DATA = datasetname;
   CLASS RESPONSE TOBACCO TRTPN (REF=first);
   MODEL RESPONSE = TOBACCO TRTPN/ DIST = BIN LINK = LOGIT;
   LSMEANS TRTPN / EXP DIFF CL;
run;
```

21.3. SECONDARY ANALYSIS OF PROPTOSIS AS A CONTINUOUS ENDPOINT

The following example code will be used to generate the MMRM ANCOVA analysis:

```
PROC MIXED DATA = datasetname;
   CLASS SUBJID TRTPN VISITNUM TOBACCO;
   MODEL CHG = BASE TOBACCO TRTPN VISIT TRTPN*VISITNUM BASE*VISITNUM/ DDFM=KR;
   REPEATED VISITNUM / SUBJECT=SUBJID TYPE=UN;
   LSMEANS TRTPN / CL;
   LSMEANS TRTPN*VISITNUM/ CL;
run;
```

where CHG = proptosis change from Baseline; BASE = Baseline proptosis; TOBACCO = tobacco use status ('Y' or 'N'); TRTPN = treatment group (1 = Placebo, 2 = Teprotumumab); VISIT = visit (6 = Week 6, 12 = Week 12, 18 = Week 18, 24 = Week 24)

21.4. EXPLORATORY ANALYSIS OF CATEGORICAL OVERALL RESPONSE

The following example code will be used to generate the proportional odds logistic model for ordered categorical data:

```
PROC LOGISTIC DATA=datasetname;
   CLASS TOBACCO TRTPN (REF=first) RESPONSE_CAT;
   MODEL RESPONSE_CAT = TOBACCO TRTPN;
   ODDSRATIO TRTPN;
run;
```

where RESPONSE_CAT = categorical overall response status at Week 24 (1 = High responder, 2 = Responder, 3 = Low responder, 4 = Non-responder); TOBACCO = tobacco use status ('Y' or 'N'); TRTPN = treatment group (1 = Placebo, 2 = Teprotumumab)