STATISTICAL ANALYSIS PLAN

Protocol No.:	SHP621-301
Protocol Title:	Oral Budesonide Suspension (OBS) in Adolescent and Adult Subjects (11 to 55 Years of Age, Inclusive) with Eosinophilic Esophagitis: A Phase 3 Randomized, Double-blind, Placebo-controlled Study
Drug:	SHP621, oral budesonide suspension (OBS), budesonide oral suspension (BOS)
Sponsor:	Shire ViroPharma, Incorporated (Shire) 300 Shire Way, Lexington, MA 02421 USA
Version No. and Date	Final Version 1.0, Date 11-March-2019

PREPARED BY

PPD, Inc.
3900 Paramount Pkwy
Morrisville, NC, 27560

Issued by: [redacted] (BIOS) I approve this document 11 Mar 2019 23:09:27 -04:00 DocuSign [redacted], Biostatistics
PPD

Date: 03/ 11 /2019

Reviewed by: [redacted] Biostatistics I approve this document 12 Mar 2019 08:07:05 -04:00 DocuSign [redacted], Biostatistics
PPD

Date: ___/___/___

Upon review of this document, including table, listing, and figure shells, the undersigned approves the statistical analysis plan. The analysis methods and data presentation are acceptable, and the table, listing, and figure production can begin.

Approved by:

[redacted]

[redacted] **Date**

[redacted], Global Clinical Development

Shire

[redacted]

[redacted] **Date**

[redacted], Global Clinical Development

Shire

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

λz	first order rate constant associated with the terminal (log-linear) portion of the curve
ACTH	adrenocorticotropic hormone
AE	adverse event
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
AUCtau	area under the curve for the defined interval between doses
BID	twice daily
BMD	bone mineral density
BMI	body mass index
BSA	Body surface area
CI	confidence interval
CL/F	apparent oral clearance
C_{max}	maximum concentration occurring at tmax
CMH	Cochran-Mantel-Haenszel
DSQ	Dysphagia Symptom Questionnaire
DXA	(DEXA) dual-energy X-ray absorptiometry
eCRF	electronic case report form
EGD	esophagogastroduodenoscopy
EoE	eosinophilic esophagitis
EoE-QoL-A	Adult Eosinophilic Esophagitis Quality of Life
EQ-5D	EuroQol
EQ-5D-3L	EuroQol-5 Dimensions 3-level
EQ-5D-Y	EuroQol 5 Dimensions Youth
ET	early termination
EREFS	EoE Endoscopic Reference Score
FAS	Full Analysis Set
HPF	high-powered field
HRQoL	health-related quality of life
hs	at bedtime
IWRS	Interactive web-based response system
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent-to-Treat

OBS	oral budesonide suspension
BOS	budesonide oral suspension
PD	Pharmacodynamics
PedsQL-EoE	Pediatric Quality of Life Inventory – EoE
PGI-S	Patient Global Impression of Severity
PK	Pharmacokinetic
PP	per-protocol
pc	after meals
qAM	every morning
SAE	serious adverse event
SAP	statistical analysis plan
SAS®	statistical analysis system
SOC	system organ class
TEAE	treatment-emergent adverse event
t_{max}	time of maximum observed concentration sampled during a dosing interval
$t_{½}$	terminal half-life
VAS	Visual Analogue Scale
V_z/F	apparent volume of distribution associated with the terminal slope
WHO	World Health Organization

1. INTRODUCTION

This statistical analysis plan (SAP) provides a technical and detailed elaboration of the SHP621-301 statistical analyses of efficacy, safety/tolerability, health related quality of life (HRQoL) and pharmacokinetic (PK) data as described in study Protocol Amendment 2, dated 26 January 2018 and SHP621-301 Protocol Amendment 2 – Administrative Change, dated 20 February 2019. Specifications for tables, figures, and listings are contained in a separate document.

2. STUDY DESIGN

2.1 General Study Design

This is a Phase 3, randomized, multicenter, double-blind, parallel-group, placebo-controlled study to evaluate the efficacy and safety of Oral Budesonide Suspension (OBS, also named BOS: Budesonide Oral Suspension) treatment administered twice daily (qAM, pc, and hs) for 12 weeks. The study will be conducted in adolescents and adults, aged 11-55 years, inclusive, with eosinophilic esophagitis (EoE) and dysphagia.

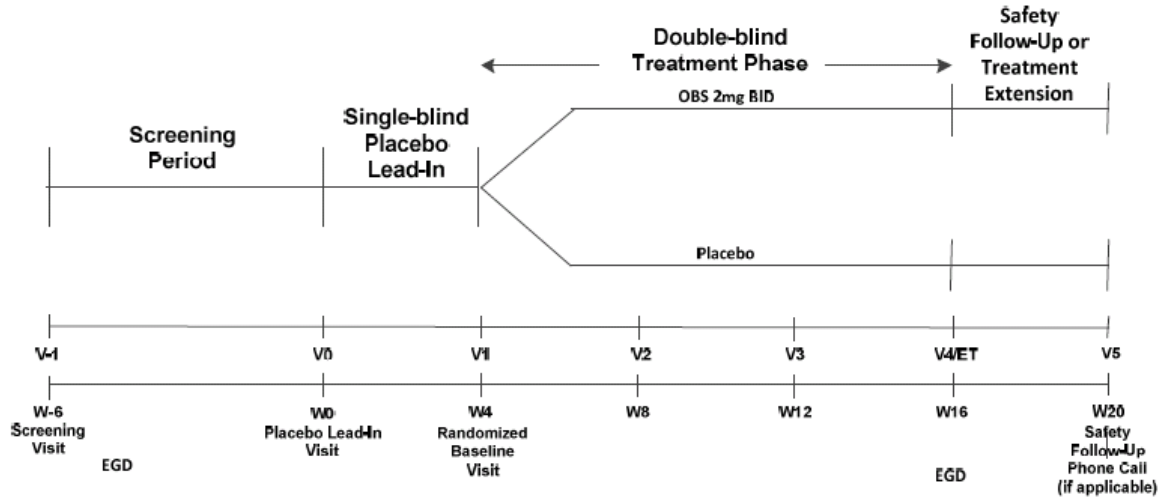
This study will consist of 3 periods: a 3- to 6-week screening period, 4-week single-blind placebo lead-in period, and a 12-week double-blind treatment period (see Figure 1). Approximately 420 subjects will be enrolled into the placebo lead-in period to allow for approximately 306 subjects to be randomized in a 2:1 ratio (approximately 204 and 102 per BOS and placebo treatment group, respectively) into the double-blind treatment period to receive either BOS 2 mg twice daily (qAM, pc, and hs) or placebo twice daily (qAM, pc, and hs). The randomization will be performed centrally and stratified by age group (2 strata total: <18 years or ≥18 years) and diet restriction for EoE or other health-related conditions (no diet restriction or any diet restriction). The stratification by age will ensure a minimum of 40 subjects in the adolescent group (11-17 years, inclusive). The stratification by age and diet will ensure balance between treatment groups for the respective stratification factors.

Subjects who sign informed consent (or assent as applicable for subjects <18 years) will be screened (Visit -1). Subjects who meet eligibility criteria at the screening visit (Visit -1) and at the placebo lead-in visit (Visit 0) will enter the 4-week placebo lead-in period to assess their ability to comply with twice daily medication administration and assess whether there is a placebo response. Upon completion of the placebo lead-in period, subjects will return for the baseline visit (Visit 1) to confirm eligibility. Eligible subjects will be randomized 2:1 into the double-blind treatment period and will be evaluated for efficacy and safety at Weeks 8, 12, and 16 (Visits 2-4). Subjects who fail to meet all eligibility criteria at Visits -1, 0, or 1 will be considered screen failures. Subjects cannot be rescreened once it is confirmed they do not meet inclusion/exclusion criteria unless the screen failure was due to a temporary condition or incomplete information at the time of consent (Visit -1) that would make rescreening at a later date appropriate (eg, concomitant medication that can be discontinued prior to rescreening; review of subject medical records provides new information with respect to the date of a prior esophageal dilation or diet change, or subject has a minor illness such as an upper respiratory or urinary tract infection). All reasons for rescreening (ie, reasons unrelated to inclusion/exclusion criteria) must be discussed and approved prospectively with the medical monitor. Subjects who discontinue will not be replaced.

At the end of the 12-week double-blind treatment period (Visit 4), subjects who complete the study will have the opportunity to enroll in the treatment extension study. These subjects will continue on the blinded assigned treatment for 2-4 weeks as part of the screening prior to enrolling into the treatment extension study.

Subjects will be required to visit the site up to 6 times over up to a 22-week period. A safety follow-up phone call will occur 4 weeks following the last dose of investigational product for subjects who discontinue prematurely during the double-blind treatment period or who do not enroll in the treatment extension study.

Figure 1 Study Design Flow Chart



Abbreviations: BID=twice daily; EGD=esophagogastroduodenoscopy; ET=end of treatment; OBS=oral budesonide suspension

2.2 Randomization

Subjects will be randomized after confirmation of study eligibility in a ratio of 2:1 via a computer-generated randomization schedule to receive BOS 2 mg twice daily (qAM, pc, and hs) or placebo twice daily (qAM,pc, and hs). The randomization will be performed centrally via Interactive Web-based Response System (IWRS) and stratified by age group (2 strata total: <18 years or ≥18 years) and diet restriction for EoE or other health-related conditions (no diet restriction or any diet restriction). The stratification by age will ensure a minimum of 40 subjects in the adolescent group (11-17 years, inclusive). The stratification by age and diet restriction will ensure balance between treatment groups for the respective stratification factors. Fixed block randomization will be used to ensure subjects are assigned to BOS or placebo in a ratio of 2:1 within strata.

2.3 Blinding

2.3.1 Blinding of Investigational Product

This is a double-blind study. Blinding was achieved by means of identical appearance for BOS and placebo. Placebo consists of all components of the investigational product solution with the exception of budesonide.

2.3.2 Blinding of Study Data

2.3.2.1 SHP621-301 Study Conduct

To protect the integrity of the study blind in both the SHP621-301 and SHP621-302 (treatment extension study of SHP621-301) studies, the post-randomization central histology and Dysphagia Symptom Questionnaire (DSQ) subject level data are segregated in separate case report forms that are not available to the blinded study team, study sites, and subjects until the final database lock of SHP621-302. Although treatment assignment will remain blinded to the blinded study team, study sites, and subjects, these subject level data could potentially identify which treatment a subject is randomized to (e.g., in individual subjects with substantial decreases in esophageal eosinophils from screening). An Unblinded Data Team (UBDT) was established to handle the processing, review, and validation of all histology and DSQ data to ensure consistency in the conduct of these activities and collection of data in electronic case report forms. The UBDT has restricted access to post-randomization histology and DSQ data during the conduct of the SHP621-301 and SHP621-302 studies, and the UBDT operates independently from the blinded study team, who are involved in study oversight or day-to-day conduct, and investigators at study sites.

2.3.2.2 SHP621-302 Study Conduct

The SHP621-302 study entitled "A Phase 3, Multicenter, Double-blind Extension Study to Evaluate Maintenance of Efficacy of Oral Budesonide Suspension (BOS) and Long-term Treatment Effect of BOS in Adolescent and Adult Subjects (11 to 55 Years of Age, Inclusive) with Eosinophilic Esophagitis (EoE)" is limited to subjects who complete the induction study and meet additional eligibility criteria. As the extension study will be ongoing at the time of the final SHP621-301 database lock, Shire will ensure that the blinded SHP621-302 study team, study sites, and subjectss remain blinded to treatment assignment and subject level central histology and DSQ data collected in both the SHP621-301 and SHP621-302 studies through the final database lock in the SHP621-302 study.

2.3.2.3 SHP621-301 Study Results Reporting

After the final SHP621-301 database lock, select team members will have restricted access to individual patient treatment assignment, and patient-level efficacy, safety and PK data from the SHP621-301 study for purposes of study reporting. A separate charter will be completed prior to the final database lock for the SHP621-301 study to define processes for communicating SHP621-301 aggregate treatment-level efficacy, safety and PK results and for restricting access to subject-level data to ensure that the SHP621-302 blinded clinical trial team (CTT) remains blinded through the final database lock for the SHP621-302 study.

2.4 Schedule of Assessments

Table 1 below presents the schematic of the study design.

Table 1 Schedule of Assessments

Procedures	Screening	Placebo Lead-in	Treatment Phase				Safety Follow-up Telephone Contact[r]
	Visit -1	Visit 0	Randomization Baseline/Visit 1	Visit 2	Visit 3	Visit 4 or ET[q]	Visit 5
Week	**-6**	**0**	**4**	**8**	**12**	**16**	**20**
Window	**≤6 weeks**	**--**	**±3 days**	**±3 days**	**±3 days**	**±3 days**	**±3 days**
Informed consent/assent	X						
Medical history review	X						
Inclusion/exclusion criteria review	X	X	X				
Vital signs[a]; height[b] and weight[c] assessment	X	X	X	X	X	X	
EGD with endoscopy score (EREFS) and biopsy[d]	X					X	
DSQ training and issue of handset	X						
Retrieval of DSQ handset						X	
DSQ completion	Once daily completion						
DSQ compliance assessment[e]		X	X	X	X	X	
EQ-5D[f]			X			X	
PedsQL-EoE (subjects 11-17 years of age, inclusive)			X			X	
EoE-QoL-A (subjects ≥18 years of age)			X			X	
PGI-S			X	X	X	X	
Physical examination	X	X	X	X	X	X	
Tanner Staging Assessment[g]	X					X	

Table 1 Schedule of Assessments

Procedures	Screening	Placebo Lead-in	Treatment Phase				Safety Follow-up Telephone Contact[r]
	Visit -1	Visit 0	Randomization Baseline/Visit 1	Visit 2	Visit 3	Visit 4 or ET[q]	Visit 5
Week	**-6**	**0**	**4**	**8**	**12**	**16**	**20**
Window	**≤6 weeks**	**--**	**±3 days**	**±3 days**	**±3 days**	**±3 days**	**±3 days**
Clinical laboratory tests[h]	X	X	X	X	X	X	
Urinalysis[i]	X	X	X	X	X	X	
Pregnancy test[j]	X	X	X	X	X	X	
Morning cortisol (target 6:00-9:00 AM)			X	X	X	X	
ACTH stimulation testing			X			X	
Blood pharmacokinetic sampling (subjects ≥18 years of age)[k]				X	X	X	
DXA Scan (subjects 11-17 years of age, inclusive)[l]		X				X	
Randomization[m]			X				
Investigational product supplied		X	X	X	X	X[l]	
Investigational product administration[o]		Twice-daily administration of investigational product					
Investigational product compliance assessment			X	X	X	X	
Concomitant medications and procedures recorded	X	X	X	X	X	X	X
Review of adverse events[p]		X	X	X	X	X	X

ACTH=adrenocorticotropic hormone; DSQ=Dysphagia Symptom Questionnaire; DXA=dual-energy X-ray absorptiometry; EGD=esophagogastroduodenoscopy; EoE-QoL-A=Adult Eosinophilic Esophagitis Quality of Life; EQ-5D=EuroQol; EQ-5D-3L=EuroQol-5 Dimension 3-level; EQ-5D-Y=EuroQol-5 Dimensions Youth; EREFS=EoE Endoscopic Reference Score; hs=at bedtime; IWRS=interactive web-based response system; PedsQL-EoE=Pediatric Quality of Life Inventory – EoE; pc=after meals; PGI-S=Patient Global Impression of Severity; qAM=every morning

[a] Vital signs will be assessed after the subject has been in a supine position for at least 5 minutes immediately prior to the assessment and will include blood pressure (systolic and diastolic), heart rate, respirations, and temperature.

Table 1 Schedule of Assessments

Procedures	Screening	Placebo Lead-in	Treatment Phase				Safety Follow-up Telephone Contact[r]
	Visit -1	Visit 0	Randomization Baseline/Visit 1	Visit 2	Visit 3	Visit 4 or ET[q]	Visit 5
Week	**-6**	**0**	**4**	**8**	**12**	**16**	**20**
Window	**≤6 weeks**	**--**	**±3 days**	**±3 days**	**±3 days**	**±3 days**	**±3 days**

[b] Height to be collected at screening visit (Visit -1) and Visit 4 for all subjects. Stadiometers are required for subjects 11-17 years of age, inclusive, and will be used at Visit -1 and 4. Height measurements for adolescent subjects (11-17 years of age, inclusive) should be measured in triplicate.

[c] Weight measurements for adolescent subjects (11-17 years, inclusive) should be measured in duplicate.

[d] Pretreatment endoscopy will be performed during the screening period (at least 2 weeks prior to placebo lead-in visit [Visit 0] to allow adequate time for processing and central review). Endoscopy should include esophageal (proximal, mid-, and/or distal), gastric, and duodenal biopsies. Final treatment evaluation EGD must include esophageal biopsies; gastric and duodenal biopsies may be done at the discretion of the investigator. Final treatment evaluation EGD should occur at or within (±) 7 days of the scheduled visit.

[e] DSQ assessment includes completion of reports on DSQ handset (ie, eligibility reports and visit confirmations).

[f] Subjects 11-17 years of age, inclusive, will complete the EQ-5D-Y; subjects ≥18 years of age will complete the EQ-5D-3L.

[g] Tanner staging assessments will be performed for all subjects aged ≥11 years until investigator confirms subject is post-puberty.

[h] Clinical laboratory tests will include the following: alkaline phosphatase, aspartate aminotransferase, alanine aminotransferase, total bilirubin, total protein, albumin, glucose, blood urea nitrogen, creatinine, sodium, potassium, chloride, calcium, carbon dioxide, hemoglobin, hematocrit, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, mean corpuscular volume, erythrocyte count, leukocyte count, neutrophils, lymphocytes, monocytes, eosinophils, basophils, and platelet count. All subjects must fast overnight prior to collection.

[i] Urinalysis parameters will include glucose, protein, specific gravity, pH, nitrite, bilirubin, ketones, hemoglobin, urobilinogen, and leukocyte esterase.

[j] The serum pregnancy test will be performed for all female subjects at the screening visit (Visit -1) and final treatment evaluation (Visit 4). Urine pregnancy tests will be performed at all other visits.

[k] Blood samples for pharmacokinetic analysis will be taken from adult subjects (aged ≥18 years). Subjects who do not participate in pharmacokinetic sampling will not be discontinued from the study and lack of participation will not be a considered protocol deviation. Blood samples for pharmacokinetic analysis can be obtained one time on any day starting 7 days after Visit 1 (Week 5) and through Visit 4 (Week 16). PK samples should be drawn, ideally, at pre-dose, 0.5 and 1 hour post-dose, and at additional post-dose time points, if feasible (2, 3, 4, 6, 8, and 12 hours post-dose).

[l] The baseline DXA scan may be performed any time during the placebo lead-in period after the subject has met all screening criteria and prior to blinded-treatment randomization. Baseline and post-treatment DXA scans should be performed using the same machine and software. Post-treatment DXA scan should occur at or within (±) 7 days of the scheduled visit.

[m] Randomization will occur via IWRS at the baseline visit (Visit 1) once the subject's eligibility for study entry is confirmed.

[n] Investigational product will be dispensed at Visit 4 to subjects who consent to enroll in the treatment extension study.

[o] Subjects will receive oral administration of 10 mL of investigational product twice daily (qAM, pc, and hs), with no ingestion of food or liquids permitted for 30 minutes after study drug administration.

[p] AE assessments at each visit and physical examination must include specific assessments for signs of glucocorticoid excess (e.g., moon facies, acne, hirsutism, mood swings,

Table 1 Schedule of Assessments

Procedures	Screening	Placebo Lead-in	Treatment Phase				Safety Follow-up Telephone Contact[r]
	Visit -1	Visit 0	Randomization Baseline/Visit 1	Visit 2	Visit 3	Visit 4 or ET[q]	Visit 5
Week	**-6**	**0**	**4**	**8**	**12**	**16**	**20**
Window	**≤6 weeks**	**--**	**±3 days**	**±3 days**	**±3 days**	**±3 days**	**±3 days**

insomnia, and depression).

[q] If subject discontinues study prematurely, the evaluations listed for Visit 4 are to be performed as completely as possible.

[r] For subjects who withdraw from the study or do not continue into treatment extension study, a safety follow-up contact by phone will be performed 4 weeks following the last dose of investigational product.

2.5 Determination of Sample Size

The co-primary efficacy endpoints are the following:

- Histologic response, defined as a peak eosinophil count of ≤6/HPF across all available esophageal levels at the final treatment period evaluation Visit 4 (Week 16)
- Dysphagia symptom response, defined as ≥30% reduction in the DSQ combined score (questions 2+3) from baseline to the final treatment period evaluation Visit 4 (Week 16)

Based on at least a 30-percentage-point reduction in DSQ score, there is an expected difference between treatment response proportions of 69% and 45% in the BOS 2 mg twice daily (qAM, pc, and hs) and placebo groups, respectively. The original protocol noted that a total of 228 subjects (152 subjects randomized to BOS and 76 subjects randomized to placebo) were required to achieve 90% power at the significance level of 0.0499 (2-sided) using a 2-group chi-square test with unequal allocation 2:1 to treatment groups (BOS 2 mg twice daily and placebo). With the specified number of subjects per treatment group, the study was originally powered at 99% assuming histological response proportions of 40% and 3% in the BOS 2 mg twice daily and placebo groups, respectively. The overall study power for the co-primary endpoints was estimated to be at least 85%. Therefore, approximately 228 (approximately 152:76 BOS and placebo subjects, respectively) were to be randomized in the study to allow for a loss of approximately 5% of subjects due to dropouts or invalid data.

In Protocol Amendment 2, in order to ensure that a sufficient number of subjects complete this study and enroll in the treatment extension study SHP621-302, up to approximately 420 subjects will be enrolled in the placebo lead-in period to allow for approximately 306 subjects to be randomized into the double blind period of this study (approximately 204:102 BOS and placebo subjects, respectively). With a total of 306 randomized subjects in this study, using the same assumptions for the co-primary endpoints and the dropout rate, the overall study power for the coprimary endpoints is estimated to be at least 95%.

The key secondary endpoint is the change in DSQ combined score (questions 2+3) from baseline to the final treatment period evaluation Week 16. Based on data from the Phase 2 study MPI 101-06 , the common standard deviation for change in DSQ combined score among similar EoE subjects following 12 week treatment of BOS is 13 points. Assuming a 5% dropout rate in each group and a common standard deviation (SD) of 13 points, a total sample size of 306 (approximately 204:102 BOS and placebo subjects, respectively) will have at least 90% power to detect a treatment difference of 5.3 points in the change of DSQ combined score from baseline to the final treatment period evaluation between treatment groups, using a 2-sided t-test at a significance level of 0.049.

2.6 Multiplicity Adjustments for Type I Error Control

The study will be deemed successful if each of the co-primary efficacy endpoints are achieved. A very minimal fraction of alpha (0.0001) was spent at the interim analysis (refer to Section 14 for interim analysis) as the study was not stopped due to the interim results. The final analysis will use 4.99% for each of the co-primary endpoints in order to preserve an overall type I error at 5% level.

A hierarchical testing procedure will be used to control the Type I error for the multiple endpoints tested at $\alpha = 0.049$. For a test to be considered statistically significant within the testing hierarchy, it must be statistically significant at the 0.0499 level, and all previous tests within the testing hierarchy must be statistically significant at the 0.0499 level. The testing hierarchy is as follows:

1. Histologic response and dysphagia symptom response (co-primary endpoints)
2. Change in DSQ combined score (questions 2+3) from baseline to the final treatment period evaluation Week 16 (key secondary endpoint)

3. OBJECTIVES

3.1 Primary Objectives

The co-primary objectives of the study are to demonstrate in a placebo-controlled study that:

- BOS induces a histologic response (eosinophilic count ≤6/HPF) in adolescent and adult subjects with EoE over a 12-week course of therapy.
- BOS reduces dysphagia, as measured by the DSQ, by at least 30% from baseline in adolescent and adult subjects with EoE over a 12-week course of therapy.

3.2 Secondary Objectives

The key secondary objective of this study is:

- BOS reduces dysphagia, as measured by the DSQ score from baseline to the final treatment period evaluation at Visit 4 (Week 16).

Additional secondary objectives of the study are:

- To assess the response of endoscopically identified esophageal features to BOS as compared to placebo as measured by the EoE Endoscopic Reference Score (EREFS)
- To explore other responding criteria based on histology and DSQ
- To assess the impact of BOS on pain, as measured by the DSQ pain score
- To evaluate the safety and tolerability of BOS over a 12-week course of therapy
- To obtain BOS pharmacokinetic data in adult subjects with EoE

3.3 Exploratory Objectives

4. SUBJECT POPULATION SETS

4.1 Screened Set

The Screened Set will consist of all subjects who have signed informed consent, and have conducted screening assessments.

4.2 Full Analysis Set

The full analysis set (FAS) will include all randomized subjects who received at least 1 dose of a double-blind investigational product. The FAS is to be used in efficacy data analyses in which subjects will be analyzed according to their assigned randomized treatment, regardless of the treatment actually received.

4.3 Safety Analysis Set

The safety analysis set will consist of all subjects who have taken at least 1 dose of any double blind investigational product. The Safety Set is to be used for all safety analyses in which subjects will be analyzed according to the treatment they received.

4.4 Randomized Set

The randomized set will include all subjects who was randomized into the double-blind treatment period.

4.5 Per-protocol Set

The per-protocol set (PP) will include all subjects in the FAS excluding subjects with major protocol deviations. The PP set will be identified prior to unbinding the treatment assignments by a team consisting of, at a minimum, a physician and a statistician from Shire.

Major protocol deviations that may lead to exclusion from the Per-protocol Set, are as follows, but not limited to:

1. Violations of inclusion and/or exclusion criteria
2. Compliance with study medication

   A subject, who has less than 70% or greater than 130% overall compliance during the double-blind treatment period with his/her assigned treatment
3. Study Treatment Administration/Dispensing

   A subject who had incorrect IP treatment kit ID assigned/dispensed or was administered/ dispensed “damaged” IP (e.g., IP with unapproved temperature excursion)
4. Prohibited concomitant medications
   - Significant use of CYP450 3A4 inhibitors (eg, ketoconazole, grapefruit juice) during the treatment period
   - Significant use of excluded medication while participating in the study

- Initiation of swallowed topical corticosteroid for EoE or systemic corticosteroid for any condition within 4 weeks of the final EGD
- Significant changes in uses of PPIs, H2 antagonists, antacids, or leukotriene inhibitors for any condition during the treatment period

5. Accidental Unblinding

   Subject unblinding occurs when a subject's randomization drug code is broken (accidentally or not per protocol) or their post-randomization pathology or DSQ results are unblinded prior to completion of double-blind treatment period

4.6 Pharmacokinetic Set and Pharmacokinetic Parameter Set

The pharmacokinetic (PK) Set will include all subjects in the safety set who received BOS treatment and provided at least one quantifiable plasma concentration of budesonide. The PK Parameter Set will include all subjects in the PK set for whom budesonide PK parameter(s) can be estimated using noncompartmenal methods.

5. SUBJECT DISPOSITION

Listing of Screen Failures (subjects who were screened in screening period, and subjects who were in placebo lead-in period but not randomized, separately) will be presented along with reasons for screen failure during screening period and placebo lead-in period.

The number of subjects included in each subject set (i.e., Screened, FAS, Safety, PP and PK) will be summarized by treatment group.

The number and percentage of subjects who completed and prematurely discontinued during the 12-week Double-blind Treatment Phase will be presented for each treatment group and overall for the FAS. Reasons for premature discontinuation from the Double-blind Treatment Phase as recorded on the study completion page of the eCRF will also be summarized (number and percentage) by treatment group. The number of subjects who consented to enroll into the extension study SHP621-302 will also be reported in the disposition summary. All subjects who prematurely discontinued during the Double-blind Treatment Phase will be listed by discontinuation reason for FAS Subjects.

Number of and percentage of subjects who prematurely discontinued during the screening and placebo lead-in period will be summarized.

Supporting listings of the discontinuation data from each study period (screening, placebo lead-in and double-blind treatment period) will be provided along with the information on adverse events (AE) that caused discontinuation (if subjects discontinued due to AE).

The number of subjects screened, entered placebo lead-in, randomized and completed will be tabulated by site. In addition, the duration of enrollment, in days, will be summarized for each site, and overall. Duration of enrollment will be calculated as (last date of contact for any subject at that site - the first date of informed consent for any subject at that site + 1).

6. PROTOCOL DEVIATIONS

The number and percentage of subjects reporting any protocol deviations and the incidence of each deviation type will be summarized by treatment group and overall for the FAS and individual responses will be presented in a data listing for all enrolled subjects. Protocol deviations will be summarized by deviation level (major or minor) and by type (ICH/GCP or Protocol Deviation) within each deviation level. Full list of deviations will be documented in a separate file.

7. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Baseline characteristics for endpoints assessed by EGD with endoscopy biopsy will be determined using assessments at the Screening Visit (Visit -1). Other demographic and baseline characteristics will be determined using assessments at the Baseline Visit 1 (Week 4) or the last observation prior to the first dose of investigational product during double-blind treatment phase, whichever is later. Descriptive summaries of demographic and baseline characteristics will be presented by treatment group for the FAS and Safety Set.

Subject demographic and baseline characteristics including age, age group (<18 years, >=18years), sex, race, ethnicity, weight, height, BMI, Tanner stage, diet stratification (no diet restriction or any diet restriction), peak eosinophil count at three esophageal levels and overall, DSQ score, histopathologic epithelial features combined total score ratio for grade and stage, and total endoscopic reference score, will be summarized by treatment group for the FAS. If there are multiple measurements of weight and height collected on the same date (eg adolescent subjects), then the average of multiple measurements will be used as baseline. Continuous variables will be summarized by descriptive statistics including number of subjects, mean, standard deviation (SD), median, Q1, Q3, minimum, and maximum. Categorical variables will be summarized by the number of subjects in each category and the percentage of subjects out of the total in the respective analysis set.

Height and weight will be used to calculate BMI and BSA using the formula below:

$$\text{BMI} = \text{weight [kg]}/(\text{height [m]})^2$$

A listing will be created to show all the demographics and baseline characteristics for all subjects in the FAS.

7.1 General Medical History

The investigator will record all clinically or medically relevant information regardless of how much time has elapsed since the date of any diagnosis into the CRF. General medical history data will be summarized by treatment group for the FAS. General medical history findings will also be listed for all subjects in FAS. Medical history will be coded using MedDRA Version 18.0.

7.2 EoE Medical History

The following information associated with EoE history will be recorded in the eCRF at the screening visit:

- Prior EoE Procedures (each with yes/no answer):
    - Endoscopy with Biopsy
    - Endoscopy with dilation

  - Gastrointestinal Tube Placement
  - Nasogastric Tube Placement
  - Diet Restriction

- Foods being eliminated from diet (yes – food is eliminated, no – food is not eliminated):
  - Egg, milk products, soy, peanuts/tree nuts, seafood, wheat, other.
- Was subject on liquid diet in last 3 months? (yes/no)
- Symptoms Related to EoE (each with yes/no answer):
  - Heartburn, chest pain, regurgitation, abdominal pain, nausea, eating problems, food impaction, weight loss, vomiting.
- Prior EoE Therapies (each with yes/no answer):
  - Diet therapy
    - Histologic Response (< 15 eos/hpf) to Diet Therapy (yes/no/not assessed)
  - Systemic (oral) corticosteroids
    - Histologic Response (< 15 eos/hpf) to Systemic (oral) corticosteroids Therapy (yes/no/not assessed)
  - Budesonide suspension
    - Histologic Response (< 15 eos/hpf) to Budesonide suspension (yes/no/not assessed)
  - Fluticasone aerosol
    - Histologic Response (< 15 eos/hpf) to Fluticasone aerosol (yes/no/not assessed)
  - Other Medication
    - Histologic Response (< 15 eos/hpf) to Other Medication (yes/no/not assessed)

EoE history will be summarized by treatment group for the FAS, and will also be listed for all the subjects in the FAS.

8. EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

8.1 Exposure to Investigational product

For all treated subjects, including those who terminated early, treatment exposure is the duration between the first dose date and last dose date:

Duration of exposure (days) = (Last dose date – First dose date +1).

In addition, total dose administered (mg) and average daily dose will be calculated and summarized by treatment group:

Actual Average Daily Dose (mg) = Sum of the doses (mg)/Duration of exposure (days).

Descriptive statistics (n, mean, SD, Q1, Q3, minimum, median, and maximum) will be presented by treatment group for duration of exposure and average daily dose.

A listing will be created by subject and treatment group and will provide exposure for the Safety Set.

8.2 Measurement of Treatment Compliance

Compliance with study medication will be assessed overall across all study visits. Subjects will be instructed to bring any remaining study medication and empty bottles to each study visit. Designated site staff will evaluate compliance by questioning the subject and evaluating the amount of study medication remaining. The subject will be questioned regarding any discrepancies.

When a bottle is returned, the site will measure the amount of investigational product remaining in the bottle in centimeters. The volume taken in mL for each bottle can be determined in centimeters based on the equation below:

Volume in mL=210 cm^3 – [22.4 cm^2 x height cm].

The percent compliance at each study visit will be determined as follows:

(volume of investigational product taken in all returned bottles)/(expected volume of investigational product to be taken).

Subjects will be considered compliant with study medication if they received no less than 70% and no more than 130% of the intended dosing as assessed at the final Treatment Period Visit.

Study drug exposure and compliance will be summarized, separately and combined, by treatment for the FAS during double blind period. In addition, study drug exposure and compliance will be summarized, separately and combined, by treatment for the Safety Set beginning with subject randomization. The overall approximate compliance rates, and total dose taken (mg) will be summarized using descriptive statistics. The total dose taken will be the approximate amount of investigational product taken (mL) as collected from eCRF data during the corresponding analysis period for FAS and the Safety Set. The n and percentage of subjects whose compliance is <70%, ≥70% and ≤130% (with sub categories of ≥70% and ≤100%, and >100% and ≤130%), or >130% will be summarized.

9. PRIOR AND CONCOMITANT MEDICATION

World Health Organization-Drug Dictionary (WHODRUG) as of March 2015 will be used to classify prior and concomitant medications by therapeutic class.

Prior medication is defined as any medication with the start and stop date prior to the date of the first dose of investigational product (Visit 1). Concomitant medication is defined as any medication with a start date prior to the date of the first dose of investigational product (Visit 1) and continuing after the first dose of investigational product (Visit 1) or with a start date between the dates of the first dose of investigational product (Visit 1) and the Safety Follow-up Contact, or with a start date between the dates of the first dose of investigational product and 31 days after last dose of SHP621-301 investigational product for subjects who do not have a Safety Follow-up Contact. Medication that starts after the first dose of SHP621-302 investigational product (i.e., Visit 1 in the SHP621-302 study) will be collected in SHP621-302 database and will not be considered as concomitant medication in SHP621-301. The definition of concomitant medication is similar to the TEAE definition in Section 11.1.

Medical/surgical procedures performed prior and during the treatment will be recorded on the eCRF, along with the date, and reason for the procedure.

Both prior/concomitant medication usage and medical/surgical procedures will be summarized by the number and proportion of subjects in each treatment group receiving each medication/procedure for the Full Analysis Set. Medications/procedures can be counted both as prior and concomitant. Multiple medication/procedure received by a subject in the same category will be counted only once.

All prior/concomitant medication and medical/surgical procedures will be listed.

10. EFFICACY ANALYSES

The primary, key secondary, and secondary efficacy analyses will be performed on the FAS and presented by treatment group. The most recent non-missing measurement (scheduled or unscheduled) collected prior to first dose of study drug administered in double-blind phase will be used as the baseline for all efficacy analyses.

10.1 Primary Efficacy Endpoint(s) and Analysis

10.1.1 Definition of Co-primary Efficacy Endpoints

The co-primary efficacy endpoints are the following:

- Histologic response, defined as a peak eosinophil count of ≤6/HPF across all available esophageal levels at the final treatment period evaluation at Visit 4 (Week 16).
- Dysphagia symptom response, defined as ≥30% reduction in the DSQ combined score (questions 2+3) from baseline to the final treatment period evaluation at Visit 4 (Week 16).

The derivation of co-primary endpoints is specified at below:

Histology Response:

An EGD with endoscopy biopsy will be performed during the study at the screening period and final treatment period evaluation Visit 4 (Week 16) or Early Termination (ET) Visit; the peak eosinophil count per high power field across all available esophageal levels at the final treatment period evaluation will be used as a primary measure of efficacy. If subject has peak eosinophil count of ≤6/HPF across all available levels then histology response variable will be set to 1 and subject will be considered as a histology responder. Otherwise, histology response will be set to 0 and subject will be treated as a non-responder. Subjects that discontinue the Double-blind Treatment Phase early and who do not have a peak eosinophil count available at Visit 4 (Week 16) will be classified as histology non-responders.

Dysphagia Symptom Response:

Subject's dysphagia symptoms will be evaluated using a DSQ ePRO device. The questionnaire will be completed each evening by subjects during the Screening Period, during the 4-week Baseline Period, and during the 12-week Treatment Period. Appendix 1 shows the daily DSQ consisting of four questions with corresponding potential point scoring in parentheses. Subjects are prompted to respond to Questions 1 first, if subjects respond positively to Question 1, then they are prompted to respond to Question 2, if subjects respond positively to Question 2, they are prompted to respond to Questions 3 and 4.

Calculations of DSQ combined score at each scheduled visit will be based on daily ePRO entries during a 2-week interval prior to each study visit. The DSQ combined score will be calculated by summing the scores of responses to questions 2 and 3 only. Responses to Questions 1 and 4 will not be included to DSQ combined score calculation.

Percent change from baseline to Visit 4 (Week 16) in DSQ combined score will be calculated as (DSQ combined score at Week 16 – DSQ combined Score at Baseline)/ DSQ combined score at Baseline*100.

Dysphagia Symptom Response (co-primary efficacy variable) will be set to 1 (responder) if subject achieved a minimum of 30% reduction in DSQ combined score between baseline and Week 16, i.e. percent change from baseline from above calculation is <=-30. Otherwise, if percent change from baseline is >-30 or subject discontinue Double-blind Treatment Phase early and DSQ combined score at Visit 4 (Week 16) is incalculable, a subject will be classified as non-responder.

10.1.2 Primary Analysis of Co-primary Efficacy Endpoints

The co-primary efficacy endpoints will be analyzed based on the FAS. DSQ combined score derivation, which defines dysphagia symptom response, will be based on the missing DSQ daily diary data handling method described in Appendix 1. Each of the co-primary efficacy endpoints is a binary response (i.e., responders vs non-responders); the co-primary endpoints will be analyzed using the Cochran-Mantel-Haenszel (CMH) test adjusting for age group (either <18 years or ≥18 years) and diet restriction for EoE or other health-related conditions (no diet restriction or any diet restriction). The adjusted odds ratio of being a responder on each of the co-primary endpoints for the BOS 2 mg twice daily group vs placebo group, associated 95% confidence interval (CI), and p-values based on CMH test will be provided. Subjects who withdraw without providing efficacy data at the final treatment period evaluation (Visit 4, Week 16) will be classified as non-responders in the primary efficacy analysis.

Additionally, the proportion of responders based on each of the co-primary endpoints for each treatment group will be summarized, and their respective 95% CI will be reported. The CMH adjusted difference in the proportion of responders between the 2 treatment groups and the corresponding 95% Newcombe CI will also be summarized. If the Newcombe CI is not available due to low response rate in either treatment group the Mantel-Haenszel confidence interval will be used. DSQ responder frequencies and percentages will be summarized for each treatment group at the final treatment period evaluation by baseline DSQ severity level subgroups (i.e., 0-21, 22-42, 43-63, 64-84).

10.1.3 Sensitivity and Other Analyses of Co-primary Efficacy Endpoints

The following sensitivity and supportive analyses will be performed for the co-primary endpoints to evaluate the robustness of the results from the primary analysis methods:

I. Each of the co-primary efficacy endpoints will be analyzed based on FAS using a logistic regression with the effects of treatment group, age group (either <18 years or ≥18 years) and diet restriction for EoE or other health-related conditions (no diet restriction or any diet restriction). The odds ratio of being a responder on each of the co-primary endpoints for the BOS 2 mg twice daily group vs placebo group and associated 95% confidence interval (CI) will be estimated from the final model. Subjects who withdraw without providing efficacy data at the final treatment period evaluation (Visit 4, Week 16) will be classified as non-responders in the primary efficacy analysis. If there are convergence problems with logistic regression, Firth logistic regression will be used.

II. CMH analyses will be repeated using the PP set.

III. CMH analyses will be repeated by considering subjects who withdraw without providing efficacy data at the final treatment period evaluation Visit 4 (Week 16) to be classified as responders.

IV. CMH analysis will be repeated based on FAS using four different missing data handling methods to assess the impact of missing DSQ daily diary data on dysphagia symptom response as follows:

Missing data handling method 1 for Dysphagia Symptom Response: Subjects who are missing 4 days or more of DSQ daily diary data in either week or 7 days in the last 2 weeks at a post-baseline visit, are treated as non-responders without applying the window shifting rule.

In the case of missing DSQ daily diary data, DSQ data for a given day will be considered missing if the diary is not reported or if the diary is reported but the answer to Question 1 is "No". If fewer than 8 reported diaries are available within the 14-day period prior to the Baseline Visit 1, the window shift rule for **baseline** DSQ calculation for the primary analysis will be performed: the most recent 8 reported diaries in a consecutive 14-day period will be used to calculate the baseline DSQ combined score. Such 14-day periods cannot be shifted for more than 7 days (not including the study visit day) when calculating DSQ combined score at Baseline Visit 1. For the 14-day period used to calculate the DSQ combined score at the post-baseline visits Visit 2 (Week 8), Visit 3 (Week 12), and Visit 4 (Week 16)/Termination Visit, 4 out of 7 days in each of the last 2 weeks are needed to calculate a DSQ combined score. The window shift rule for **post-baseline** data, which is described in Appendix 1 for the primary analysis will not be applied for this sensitivity analysis. Subjects who have missing data for at least 4 days in either of the last 2 weeks, or have missing data for at least 7 days in the last 2 weeks (ie, the 14-day period), will be treated as non-responder of Dysphagia Symptom Response, and the corresponding DSQ combined score at the visit will be incalculable.

Missing data handling method 2 for Dysphagia Symptom Response: Window shift rule is applied for post-baseline data (14-day periods cannot be shifted for more than 7 days at every visit).

In the case of missing DSQ daily diary data, DSQ data for a given day will be considered missing if the diary is not reported or if the diary is reported but the answer to Question 1 is "No". If fewer than 8 reported diaries are available within the 14-day period prior to the Baseline Visit 1, Visit 2 (Week 8), Visit 3 (Week 12) or Visit 4 (Week 16), the window shift rule for **baseline** or **Week 16** DSQ calculation in the primary analysis will be performed (see Appendix 1), the most recent 8 reported diaries in a consecutive 14-day period would be used to calculate the DSQ combined score. Such 14-day periods cannot be shifted for more than 7 days (not including the study visit day) when calculating the DSQ combined score at Baseline Visit 1 or Visit 4 (Week 16). If the DSQ combined score at Visit 1 or Visit 4 (Week 16) is incalculable, a subject will be classified as a non-responder.

Missing data handling method 3 for Dysphagia Symptom Response: Worst case imputation for DSQ combined score when the answer to Question 1 is "No".

For days the response to diary Question 1 is "No" during the 14-day period prior to the Baseline Visit 1 and the Final Treatment Period Evaluation Visit 4, the missing sum of points from question 2+3 for those days will be imputed with the worst (maximum) value of the sum of points from questions 2+3 from non-missing days during that 14-day period, After imputing the sum of points from questions 2+3 with the worst scores when the answer to Question 1 is "No", the above **missing data handling method 1** will be applied to **post-baseline data**, for subjects with missing data for at least 4 days in either of last 2 weeks prior to Visit 2 (Week 8), Visit 3 (Week 12) or Visit 4 (Week 16), or for subjects with missing data for at least 7 days in the last 2 weeks (ie, the 14-day period) before these visits being treated as non-responders. In such cases, the corresponding DSQ combined score at Week 16 will be incalculable.

For the 14-day period used to calculate the DSQ score at the **Baseline** Visit 1, the window shift rule for baseline DSQ calculation in the primary analysis will be performed: if fewer than 8 reported diaries are available within the 14-day period prior to the Baseline Visit 1, the window shift rule for baseline DSQ calculation in the primary analysis will be performed to calculate baseline DSQ score, and such 14-day periods cannot be shifted for more than 7 days. No imputation will be performed when calculating DSQ score at Baseline Visit 1.

Missing data handling method 4 for Dysphagia Symptom Response: **Best case imputation for DSQ combined score when answer to Question 1 is "No".**

For days the response to diary Question 1 is "No" during the 14-day period prior to the Baseline Visit 1 and the Final Treatment Period Evaluation Visit 4, the missing sum of points from question 2+3 for those days will be imputed with the best (minimum) value of the sum of points from questions 2+3 from non-missing days during that 14-day period, After imputing the sum of points from questions 2+3 with the best scores when the answer to Question 1 is "No", the above **missing data handling method 1** will be applied to **post-baseline data**, for subjects with missing data for at least 4 days in either of last 2 weeks prior to Visit 2 (Week 8), Visit 3 (Week 12) or Visit 4 (Week 16), or for subjects with missing data for at least 7 days in the last 2 weeks (ie, the 14-day period) before these visits being treated as non-responders. In such cases, the corresponding DSQ combined score at Week 16 will be incalculable.

For the 14-day period used to calculate the DSQ score at the **Baseline** Visit 1, the window shift rule for baseline DSQ calculation in the primary analysis will be performed: if fewer than 8 reported diaries are available within the 14-day period prior to the Baseline Visit 1, the window shift rule for baseline DSQ calculation in the primary analysis will be performed to calculate baseline DSQ score, and such 14-day periods cannot be shifted for more than 7 days. No imputation will be performed when calculating DSQ score at Baseline Visit 1.

V. Distribution-based imputation.

If at least 5% subjects or more have missing responses for either of the co-primary efficacy endpoints in either treatment group, the missing responses for co-primary efficacy endpoints will be imputed using distribution-based method.

The subjects with missing responses of co-primary efficacy endpoints will be assigned randomly according to the distribution of responders with available data for each of the co-primary endpoints (i.e., those with non-missing data) across the 2 treatment groups by strata (see Table 2).

Table 2 Percentage of Responders for All Available Data (i.e., Non-missing Data) by Strata

Strata 1	Strata 2	No	Yes
Diet	<18 years	X00%	X01%
Diet	≥18 years	Y00%	Y01%
No Diet	<18 years	X10%	X11%
No Diet	≥18 years	Y10%	Y11%

For instance, if there are N subjects with missing data in strata (age <18 years and Diet), then X00%*N subjects will be randomly assigned as non-responders and X01%*N subjects will be randomly assigned as responders.

Conversely, if there are M subjects with missing data in strata (age ≥18 years and diet), the Y00%*M subjects will be randomly assigned as non-responders and Y01%*M subjects will be randomly assigned as responders. Once the missing data are imputed a logistic regression model will be fitted with the binary outcome (responders vs non-responders) as the dependent variable and the age group, diet group, and the treatment group as independent variables. The odds ratio of being a responder on each of the co-primary endpoints for the BOS 2 mg twice daily group vs placebo group and associated 95% confidence interval (CI) will be estimated from the final model.

VI. CMH analysis based on multiple imputations (MIs).

If at least 5% subjects or more have missing responses for either of the co-primary efficacy endpoints in either treatment group, the missing responses for co-primary efficacy endpoints will be imputed with multiple imputations (MIs) by using SAS procedures PROC MI and PROC MIANALYZE. Multiple imputations (MIs) wiht the CMH analysis will be handled as follows:

- Impute missing values using the logistic regression model from effects of treatment group, age group (either <18 years or ≥18 years) and diet restriction for EoE or other health-related conditions (no diet restriction or any diet restriction).
- Repeated the process K (K=50) times, using the procedure described above to form K imputed completed datasets.

- CMH analysis will be conducted to each imputed dataset adjusting for age group (either <18 years or ≥18 years) and diet restriction for EoE or other health-related conditions (no diet restriction or any diet restriction). The values of the general association test statistic from the CMH analysis will be transformed using the Wilson-Hilferty transformation to create a normally-distributed Z scale statistic.

$$Z = \frac{(CMH)^{\frac{1}{3}} - \frac{7}{9}}{(\frac{2}{9})^{\frac{1}{2}}}$$

- Mean of the estimated co-primary endpoints from K imputations, and the resulting transformed Z values will be combined using PROC MIANALYZE in SAS to yield the corresponding p-values.

10.1.4 Subgroup Analyses of Co-primary Efficacy Endpoints

For the FAS, subgroup analyses of the co-primary efficacy endpoints will be performed in a manner similar to that of the primary analysis within each of the following subgroups:

- Age as stratification factor (<18, and ≥18 years)
- Diet restriction for EoE or other health-related conditions as stratification factor (no diet restriction or any diet restriction)

For the subgroup analysis on age, co-primary endpoints will be analyzed using the CMH test adjusting for diet restriction for EoE or other health-related conditions (no diet restriction or any diet restriction). For the subgroup analysis on diet restriction for EoE or other health-related conditions, co-primary endpoints will be analyzed using the CMH test adjusting for age group (either <18 years or ≥18 years).

10.2 Key Secondary Efficacy Endpoint(s) and Analysis

10.2.1 Definition of Key Secondary Efficacy Endpoint

The key secondary efficacy endpoint is defined as the change in DSQ combined score (questions 2+3) from baseline to the final treatment period evaluation Week 16. The key secondary efficacy endpoints will be analyzed based on the FAS.

10.2.2 Primary Analysis of Key Secondary Efficacy Endpoint

The same missing data approach described in Appendix 1 for the primary analysis of dysphagia symptom response (one of the co-primary endpoints) will be applied to the primary analysis of key secondary efficacy endpoint. The change from baseline DSQ combined score at the final treatment period evaluation Week 16 will be analyzed using an analysis of covariance (ANCOVA) model with treatment group (BOS vs Placebo) and age group (<18, ≥18 years of age) as factors and the baseline DSQ score as a continuous covariate. In addition, descriptive summary statistics, frequency distribution table and histogram will be reported for DSQ combined score at baseline, DSQ combined score at final treatment period evaluation Week 16, and the change of DSQ combined scores from baseline to Week 16.

If the DSQ combined score at Week 16 for a subject is incalculable after missing data handling, the change of DSQ score for this subject will be considered as missing in ANCOVA analyses.

10.2.3 Sensitivity Analyses of Key Secondary Efficacy Endpoint

To assess the impact of missing DSQ daily diary data on the change in DSQ combined score (questions 2+3) from baseline to the final treatment period evaluation Week 16, ANCOVA model analysis for the primary analysis of key secondary efficacy endpoint will be repeated based on FAS using four missing data handling methods described in sensitivity analyses IV of Section 10.1.3, if the DSQ combined score at Week 16 for a subject is incalculable after missing data handling method 1 or method 2, the change of DSQ combined score for this subject will be considered as missing in ANCOVA analyses.

If at least 10% subjects or more are missing DSQ combined score at Week 16 in either treatment group, the effect of missing patterns on treatment effect (i.e., missing due to no solid food taken or reasons to drop out early) will be assessed and identified, and an overall treatment effect will be obtained from the pattern mixture analysis.

10.2.4 Subgroup Analysis of Key Secondary Efficacy Endpoint

For the FAS, subgroup analysis of the key secondary efficacy endpoints will be performed in a manner similar to that of the primary analysis within the following age subgroups:

- Age as stratification factor (<18, and ≥18 years)

Key secondary efficacy endpoint will be analyzed using the ANCOVA model analysis with treatment group as factors and the baseline DSQ combined score as a continuous covariate.

10.3 Other Secondary Efficacy Endpoint(s) and Analysis

The additional secondary efficacy endpoints are the following:

- Change in total endoscopy score, as measured by the EREFS classification, from baseline to the final treatment period evaluation Week 16 (continuous)
- Peak eosinophil count <15/HPF across all available esophagus levels at the final treatment period evaluation Week 16 (binary)
- Peak eosinophil count ≤1/HPF across all available esophagus levels at the final treatment period evaluation Week 16 (binary)
- Change from baseline in the peak eosinophil count to the final treatment period evaluation Week 16 for each available esophageal level (proximal, mid-, and distal) (continuous)
- Change from baseline in the histopathologic epithelial features combined total score (grade and stage) to the final treatment period evaluation Week 16 (continuous)
- Dysphagia symptom response (binary response), defined as a ≥50% reduction in the DSQ combined score (questions 2+3), from baseline to the final treatment period evaluation Week 16 (binary)
- Change from baseline in the DSQ combined score (questions 2+3) over time including post baseline visits (continuous)

- Cumulative distribution function curves for the change and the percent change in the DSQ combined score from baseline to the final treatment period evaluation Week 16 (continuous)
- Overall binary response I, defined as a reduction in the DSQ combined score of ≥30% from baseline to the final treatment period evaluation Week 16 and a peak eosinophil count of ≤6/HPF across all esophageal levels at the final treatment period evaluation Week 16 (binary)
- Overall binary response II, defined as a reduction in the DSQ combined score of ≥50% from baseline to the final treatment period evaluation Week 16 and a peak eosinophil count of ≤6/HPF across all esophageal levels at the final treatment period evaluation Week 16 (binary)
- Change in the DSQ + pain score (questions 2+3+4) from baseline to the final treatment period evaluation Week 16 (continuous)
- Change in the DSQ pain score (question 4) from baseline to the final treatment period evaluation Week 16 (continuous)

The binary response endpoints will be analyzed using the same logistic model as the co-primary efficacy endpoints. Subjects who prematurely discontinued without efficacy evaluations at final treatment visit will be considered as non-responders for the analysis. If there are convergence problems with logistic regression, Firth logistic regression will be used.

Refer to Appendix 5 for the analysis for histopathologic epithelial features. Other continuous endpoints will be analyzed as a change from baseline using an ANCOVA model that includes treatment group and age group as factors and baseline score as a covariate. In addition, descriptive summary statistics will be reported for DSQ combined score at baseline, at each post baseline assessment visit, and change from baseline at post baseline assessment visit.

The analyses for all secondary efficacy endpoints (including the key secondary efficacy endpoint) will be carried out on FAS using 2-sided tests at the 5% level of significance. For each of the secondary efficacy endpoints, the treatment difference, corresponding 95% CI for the difference, and treatment comparison p-value for testing the null hypothesis of zero treatment effect based on the final statistical model (i.e., either logistic regression model or ANCOVA model) will be provided.

Refer to Appendix 1 to derive the DSQ combined score, DSQ + pain score, DSQ pain score, DSQ question 2 score and DSQ question 3 score at each scheduled visit using the same rules applied for the primary analysis of DSQ response. Descriptive summary statistics for each score at baseline and every post-baseline visit will be reported. Descriptive summary statistics, frequency distribution table and histogram for the percent change of DSQ combined score, change of DSQ + pain score, change of DSQ pain score, change of DSQ question 2 score and change of DSQ question 3 score from baseline to the final treatment period Week 16.

Change from baseline in the DSQ combined score (questions 2+3) overtime will be summarized descriptively for each treatment group by baseline DSQ severity level subgroups (i.e., 0-21, 22-42, 43-63, 64-84).

The total endoscopy score is the sum of endoscopy scores of proximal and distal locations from the following 5 major feature categories: 1) exudates or plaques (grade 0–2); 2) fixed esophageal rings (grade 0–3); 3) edema (grade 0–2); 4) furrows (grade 0–2); and 5) strictures (grade 0–1).The 10^{th}, the 25^{th}, the 50^{th} (Median), the 75^{th}, and the 90^{th} percentile change scores and sample sizes will be included in the cumulative distribution function curves. Besides the cumulative distribution curves specified above, the following additional secondary efficacy figures will be plotted, 95% Confidence interval band will also be plotted for all cumulative distribution curves:

- The proportion of subjects who have a histology response as a peak eosinophil count of ≤6/HPF, <15/HPF and ≤1/HPF across all available esophageal levels at the final treatment period evaluation at Week 16 will be plotted by treatment group.
- The proportion of subjects who have a reduction of ≥30% and 50% of DSQ combined score from baseline will be plotted across visits (Week 4 (baseline), Weeks 8, 12, and 16) for each treatment group.
- The proportion of subjects who have an overall response, defined as histology response as a peak eosinophil count of ≤6/HPF across all available esophageal levels at the final treatment period evaluation at Week 16, and a reduction in the DSQ combined score of ≥30% from baseline to the final treatment period of Week 16, will be plotted by treatment group.
- The proportion of subjects who have an overall response, defined as histology response as a peak eosinophil count of ≤6/HPF across all available esophageal levels at the final treatment period evaluation at Week 16, and a reduction in the DSQ combined score of ≥50% from baseline to the final treatment period of Week 16, will be plotted by treatment group.
- The mean DSQ combined score for each treatment group and standard deviations by visit (Week 4 (baseline), Weeks 8, 12, and 16) will be plotted.
- A scatter plot with the maximum overall peak eosinophil count from the esophageal biopsies (x-axis) vs. DSQ combined score at the final treatment period evaluation Week 16. Different symbols will be used for each treatment group.
- A scatter plot with the change from baseline in maximum overall peak eosinophil count from the esophageal biopsies (x-axis) vs. change from baseline in DSQ combined score at the final treatment period evaluation Week 16. Different symbols will be used for each treatment group.
- A scatter plot with the percent change from baseline in maximum overall peak eosinophil count from the esophageal biopsies (x-axis) vs. the percent change from baseline in DSQ combined score at the final treatment period evaluation Week 16. Different symbols will be used for each treatment group.

10.4 Exploratory Efficacy Endpoint(s) and Analyses

[redacted]

11. SAFETY ANALYSES

Safety analyses will be performed by treatment group using the Safety Set. Safety variables include AEs, physical examinations, stadiometry, vital signs (temperature, systolic and diastolic blood pressure, heart rate, and respiratory rate), weight and height assessments, DXA scans for BMD and body composition measurements (for adolescents aged 11-17 years, inclusive), clinical laboratory tests (hematology, chemistry, urinalysis; serum pregnancy test, if appropriate), and ACTH stimulation tests. To account for the effects of puberty in adolescent subjects (11-17 years of age, inclusive), BMD z-scores will be adjusted for height z-score. Continuous safety parameters will be descriptively summarized by treatment group at baseline and for each post-baseline visit. The last assessment prior to the first dose of investigational product will be used as baseline for all safety analyses.

11.1 Adverse Events

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), Version 18.0 or newer.

All AEs will be collected from the time of the informed consent through the completion of SHP621-301 post-treatment follow-up as described below. For subjects who do not consent to participate in the SHP621-302 study, AEs will be collected through the Safety Follow-up Contact. For eligible subjects who consent to participate in the SHP621-302 extension study, AEs will be collected in the SHP621-301 clinical database up to the date of the first dose of SHP621-302 investigational product (i.e., enrollment Visit 1 in the SHP621-302 study) or the Safety Follow-up Contact for subjects who screen fail prior to enrolling in extension study SHP621-302. AEs that have a start date after the first dose of SHP621-302 investigational product will be collected in SHP621-302 database.

TEAEs are defined as AEs that start or deteriorate on or after the first dose of double-blind investigational product (Visit 1) and through the Safety Follow-up Contact, or 31 days after the last dose of SHP621-301 investigational product for subjects who do not have a Safety Follow-up Contact. This TEAE definition is a rule for programming TEAEs and will be applied only to the SHP621-301 database; and thus, events that occur after the first dose of SHP621-302 investigational product (i.e., Visit 1 in the SHP621-302 study) will not be considered as TEAEs in SHP621-301.

However, for any subjects who die during the study (i.e., the date of death is between the date of first dose of investigational product and the date of study discontinuation entered by the site, inclusive), all AEs (including those resulting in death) that occur during the study will be considered as TEAEs irrespective of the last dose and will be included in the TEAE summaries.

An overall summary of the number of subjects with TEAEs will be presented, including number and percentage of subjects with any TEAEs, serious TEAEs, severe TEAEs, life-threatening TEAEs, TEAEs related to investigational product, TEAEs related to EoE, deaths and hospitalizations due to TEAEs, TEAEs leading to discontinuation of investigational product and TEAEs leading to study discontinuation.

The number and percentage of subjects reporting TEAEs in each treatment group will be tabulated by system organ class (SOC) and preferred term, by SOC, preferred term, and maximum severity. Serious TEAEs, TEAEs considered related to investigational product, and TEAEs related to EoE will also be summarized by SOC and preferred term. If more than 1 AE occurs with the same preferred term for the same subject, then the subject will be counted only once for that preferred term using the most severe and most related occurrence for the summarization by severity and by relationship to investigational product.

The incidence of common TEAEs (≥2% of subjects in any treatment group) will be summarized by preferred term. Serious TEAEs, TEAEs leading to discontinuation of investigational product and TEAEs leading to discontinuation from the study will be summarized by SOC, preferred term and treatment group. Deaths will be summarized by preferred term and treatment group.

11.2 Clinical Laboratory Variables

Descriptive statistics for clinical laboratory values (in US conventional units) and changes from baseline at each assessment time point as well as shift tables from baseline to each visit for quantitative variables will be presented by treatment group for the following clinical laboratory variables. All laboratory data will also be listed.

Hematology hemoglobin, hematocrit, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, mean corpuscular volume, mean platelet volume, erythrocyte count, erythrocyte distribution width, leukocyte count, neutrophils, lymphocytes, monocytes, eosinophils, basophils, and platelet count.

Biochemistry alkaline phosphatase, aspartate aminotransferase, alanine aminotransferase, total bilirubin, total protein, albumin, glucose, blood urea nitrogen, creatinine, sodium, potassium, chloride, calcium, carbon dioxide.

Urinalysis glucose, protein, specific gravity, pH, nitrite, bilirubin, ketones, hemoglobin, urobilinogen, and leukocyte esterase.

Other tests serum pregnancy, urine pregnancy, morning cortisol (6:00-9:00 AM collection), ACTH stimulation testing (serum cortisol collections at 30 and 60 minutes after synthetic ACTH injection in addition to baseline collection at 6:00-9:00 AM).

Pregnancy test results will only be listed. A serum β-hCG pregnancy test is performed on all female subjects at the screening visit (Visit -1) and the final treatment evaluation visit (Visit 4) or ET visit. A urine pregnancy test is performed on all female subjects at the placebo lead-in visit (Visit 0), baseline visit (Visit 1), Visit 2, and Visit 3 or if pregnancy is suspected.

ACTH stimulation testing will be performed by measuring the levels of cortisol in the blood following the injection of a synthetic form of ACTH. The type of synthetic and route of administration will be per local lab discretion. Blood samples will be collected just prior to and approximately 30 and 60 minutes following the injection at Baseline Visit 1 (Week 4) and Week 16.

Morning unstimulated cortisol testing results will be descriptively summarized by treatment group and visit. For ACTH stimulation testing results, the number and percentage of subjects with the highest of the two cortisol values at the 30 and 60 minute time points that meet a threshold criterion event of ≤18 mcg/dL will be summarized by treatment group at Baseline Visit 1 (Week 4) and Week 16, and worsening of the shift from Baseline Visit 1 to Week 16 will be presented. ACTH stimulation testing results will also be presented with a higher threshold criterion of ≤ 20 mcg/dL, i.e., the number and percentage of subjects with the highest of the 2 cortisol values at the 30 and 60 minutes time points of ≤ 20 mcg/dL.

11.3 Vital Signs

Descriptive statistics for vital signs (systolic and diastolic blood pressure, heart rate, respirations, temperature and weight) and their changes from baseline at each post-baseline visit and at the end of study will be presented by treatment group. Weight is collected at every visit for all subjects. For adolescent subjects 11-17 years of age, inclusive, two weight measurements per visit will be recorded. The average weight per visit will be used in summary tables for these subjects.

A separate summary of height will be provided by visit and treatment group. Height will be collected at screening visit (Visit -1) and Visit 4 for all subjects. For adolescent subjects 11-17 years of age, inclusive, stadiometers will be used to measure height, and three height measurements per visit will be recorded. The average height per visit will be used in summary tables for these subjects.

All vital signs data will also be listed.

11.4 Electrocardiogram (ECG)

Not applicable since ECG assessments are not planned for this study.

11.5 Other Safety Variables

Physical Examination

Physical examination assessments at each visit will include specific assessments for signs of glucocorticoid excess (e.g., moon faces, acne, hirsutism, mood swings, insomnia, and depression). Physical examination at the screening visit (Visit -1) will also include Tanner Staging Assessments for subjects <18 years of age. The number and percentage of subjects reporting symptoms described above will be presented by visit and treatment group.

Dual-energy X-ray Absorptiometry for Bone Mineral Density

DXA (also referred to as DEXA) scans for determination of BMD and body composition will be performed in subjects aged 11-17 years, inclusive, at baseline and during last treatment Visit 4 (Week 16). The sites for DXA measurement will be the lumber spine (L1-L4 preferred) and total body less head. To account for the effects of puberty in adolescent subjects (11-17 years of age, inclusive), BMD z-scores will be adjusted for height z-scores.

Mean Z-scores at scheduled assessments and mean within-subjects changes from baseline at Visit 4 (Week 16) will be calculated for the DXA Z-scores; corresponding two-sided 95% confidence intervals will be provided.

12. CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Pharmacokinetic analysis will be performed by PPD (Richmond, VA) under the supervision of Shire.

12.1 Pharmacokinetic Methods

The following sections only pertain to the noncompartmental PK analysis of data collected in the SHP621-301 study. A population PK analysis will be performed separately using the pooled PK data from this study as well as historical studies; the population PK report will be prepared separately.

All PK concentrations of budesonide will be presented using the PK set; summaries and analyses of the PK parameter data will be based on the PK Parameter set.

12.1.1 Concentration Data

Blood samples will be drawn from adult subjecs (≥18 years of age) during this study for the determination of plasma concentrations of budesonide. Serial blood samples will be collected once during the treatment phase for PK analysis at predose, and at 0.5, and 1 hour post-dose in all subjects. Additional blood samples will be collected at 2, 3, 4, 6, 8, and 12 hours post-dose in subjects with intensive PK sampling. Actual sampling times post-dose will be considered protocol deviations where the sample deviated from nominal collection time by more than ±5 minutes within the first 4 hours and ±15 minutes after 4 hours post-dose. Samples collected outside these windows will be flagged in data listings. Plasma concentrations of budesonide will be determined using a validated bioanalytical method.

Individual budesonide plasma concentrations will be listed for all subjects by subject, visit, and time based on the PK set and summarized by time (overall and by visit [week]) based on the PK set with the following descriptive statistics: number of observations (n), arithmetic mean, standard deviation (SD), coefficient of variation (CV%), median, minimum, maximum, geometric mean, and geometric CV.

Budesonide plasma concentrations will also be summarized based on the PK Parameter set. The mean and individual budesonide plasma concentration versus time profiles will be presented in figures on both linear and semi-logarithmic scales based on PK Parameter set. Mean (±SD) budesonide plasma concentration versus time profiles will be presented for the PK Parameter set using nominal time and all sampling times considered to be protocol deviations will be excluded from summaries and mean plots. Individual budesonide plasma concentration versus time profiles will be presented using actual time based on the PK Parameter set.

12.1.2 Handling BLQ Values

The following procedures will be used for plasma concentrations that are below the lower limit of quantification (LLOQ):

- Plasma samples that are below the limit of quantification (BLQ) will be reported as zero in the data listings.

- Samples that are BLQ will be treated as zero in the calculation of summary statistics (e.g., mean, SD, etc.) for the plasma concentrations at individual time points. Geometric mean will be set to missing where zero values exist.
- Mean concentrations will be reported as zero if all values are BLQ, and no descriptive statistics will be reported. If the calculated mean (± SD) concentration is less than the LLOQ, the value will be reported as calculated. The mean values derived using these conventions will be used to create the mean plasma concentration versus time plots.
- For calculation of area under the plasma concentration curve (AUC), BLQ values will be set equal to zero in the dataset loaded into Phoenix® WinNonlin® (Certara USA, Inc, Princeton, NJ) for PK analysis. WinNonlin® uses the zero values that occur before the first time point with a concentration greater than LLOQ. Values that are BLQ after the first measurable concentration will be set to "missing" in the dataset loaded into WinNonlin®.
- Missing values will not be imputed.

12.1.3 Pharmacokinetic Parameters

Pharmacokinetic parameters will be determined from the plasma concentration-time data for budesonide by non-compartmental analysis using Phoenix® WinNonlin® (Certara, Princeton, NJ) Version 6.4 or higher for all subjects included in the PK Parameter set (i.e. subjects who provide at least 4 post-dose PK samples). All calculations will be based on actual sampling times, and all PK samples will be included in the PK analysis (including those with actual times outside the predefined collection interval and considered to be protocol deviations).

The pharmacokinetic parameters will include, but may not be limited to, those listed in Table 3, subject to the quality of the PK data.

Table 3 Pharmacokinetic Parameters

Parameter	Definition
$AUC_{0\text{-}tau}$	Area under the curve for the defined interval between doses (12 hours), calculated using the linear-up/log-down trapezoidal rule In this method, the linear trapezoidal method of calculation is used for each incremental trapezoid and the log trapezoidal rule for each decremental trapezoid
C_{max}	Maximum concentration
t_{max}	Time of maximum observed concentration sampled during a dosing interval
λz	First order rate constant associated with the terminal (log-linear) portion of the curve
$t_{½}$	Terminal half-life
CL/F	Apparent oral clearance
V_z/F	Apparent volume of distribution associated with the terminal slope

PK parameters will be summarized based on the PK Parameter set. The PK parameters for budesonide will be summarized with descriptive statistics (n, arithmetic mean, SD, CV%, median, maximum, minimum, geometric mean, CV% of geometric mean, and 95% CI of geometric mean) will be determined for all pharmacokinetic parameters overall and by visit (week).

Population PK analysis will be performed for all subjects who provide PK samples (limited or intensive samples). Full details of the population PK analysis are documented separately.

12.2 Pharmacodynamic Methods

Not applicable

12.3 Statistical Analysis of Pharmacodynamic Data

Not applicable

13. OTHER ANALYSES

13.1 Quality of Life Analyses

Adult Eosinophilic Esophagitis Quality of Life Questionnaire

The EoE-QoL-A will be administered in subjects ≥18 years of age at baseline Visit 1 and final Visit 4. The EoE-QoL-A is a disease-specific measure of HRQoL in adult subjects (≥18 years of age) with EoE (Taft et al., 2011). The EoE-QoL-A consists of a 30-item test with 5 subscales: eating/diet impact, social impact, emotional impact, disease anxiety, and choking anxiety. For each of the 30 questions subjects will be asked to provide a response on a 5-point scale: not at all, slightly, moderately, quite a bit and extremely.

Results of the EoE-QoL-A questionnaire will be summarized for each of the 5 scales/30 questions by presenting number of subjects and percentages by treatment group and visit.

Summary of change from baseline to the final treatment period evaluation Visit 4 (Week 16) in EoE-QoL-A total score and each of subscales will also be presented. Refer to Appendix 2 for derivation of EoE-QoL-A total score.

Pediatric Quality of Life – EoE Questionnaire

The PedsQL-EoE questionnaire will be completed by subjects 11-17 years of age, inclusive, and their parent or legal guardian, at baseline Visit 1 (Week 4) and final Visit 4 (Week 16).

The PedsQL-EoE is a modular, disease-specific instrument designed to measure quality of life in children and adolescents (2-18 years of age) with EoE (Franciosi, 2013). The PedsQL-EoE module consists of 35 items for children and teenagers encompassing the following 7 scales: 1) Symptoms I (6 items; chest/throat/stomach pain and nausea/vomiting), 2) Symptoms II (4 items; trouble swallowing), 3) Treatment (5 items; treatment barriers), 4) Worry (6 items; worries about treatment and disease), 5) Communication (5 items; communication with others about EoE), 6) Food and Eating (4 items; food and eating allergies and limitations), and 7) Food Feelings (3 items; emotions associated with food allergies).

Results of the PedsQL-EoE report (adolescents and parents separately) questionnaire will be summarized for each of the 7 scales/questions by presenting number of subjects and percentages by treatment group and visit.

Summary of change from baseline to the final treatment period evaluation Visit 4 (Week 16) in PedsQL-EoE total score (adolescents and parents separately) and each of subscales will also be presented. Refer to Appendix 3 for derivation of PedsQL-EoE total score.

EuroQol-5 Dimensions 3-level Questionnaire

The EuroQol-5D Dimensions 3-level (EQ-5D-3L) will be for subjects ≥18 years (EuroQol, 1990; Brooks, 1996).

The EuroQol-5 Dimensions Youth (EQ-5D-Y) will be for subjects 11-17 years of age, inclusive (Willie et al., 2010). The EQ-5D-3L and EQ-5D-Y will be administered during the study at baseline Visit 1 (Week 4) and final Visit 4 (Week 16).

EQ-5D-3L consists of 2 pages – the EQ-5D descriptive system and the EQ visual analogue scale (EQ VAS). The EQ-5D-3L descriptive system comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort, anxiety/depression, each of which can take 1 of 3 responses. The responses record 3 levels of severity within a particular dimension – (1) no problems, (2) some problems, (3) extreme problems. The EQ VAS records the respondent's self-rated health on a vertical, visual analogue scale where the endpoints are labelled 'Best imaginable health state' (100) and 'Worst imaginable health state' (0).

The EQ-5D-Y was developed to assess quality of life in children. As this version was intended to be comparable with the adult version EQ-5D-3L, the wording and layout of the EQ-5D-3L was modified to ensure relevance and clarity for the cognitive developmental stage of children. The measure requires children to self-report on five dimensions of health, namely; mobility (walking around), taking care of myself, doing usual activities, having pain or discomfort and feeling worried, sad or unhappy. The child can choose between three levels of severity (1) no, (2) some or (3) a lot of problems, which result in a health profile. Additionally, a Visual Analogue Scale (VAS) allows the child to subjectively rate their overall quality of life on a graduated scale, with 0 indicating worst health state imaginable and 100 indicating best health.

Results of both EQ-5D-3L and EQ-5D-Y descriptive system instruments will be summarized separately on each question by presenting number of subjects and percentages by treatment group and visit. VAS at baseline and final treatment period evaluation Visit 4 (Week 16), and change from baseline to the final treatment period will be summarized using the following descriptive statistics: the number of subjects, mean, standard deviation, standard error, minimum value, 25th percentile, median, 75th percentile, and maximum value.

Descriptive summary of EQ-5D-3L composite index score at baseline and final treatment period evaluation Visit 4 (Week 16), and change from baseline to the final treatment period evaluation Visit 4 (Week 16) will also be presented. Refer to Appendix 4 for derivation of EQ-5D-3L composite index score.

14. INTERIM ANALYSIS

An unblinded interim analyses of the co-primary efficacy endpoints was performed by external independent unblinded statisticians and programmers at the Contract Research Organization, PPD, after 114 randomized subjects had either completed the study or withdrew prematurely from the study. The total of 114 randomized subjects represents 50% of the projected total randomization of 228 subjects based on protocol amendment 1 (i.e. information time t=0.5). No consideration was given to stop the study early at the time of interim analyses. A very minimal fraction of alpha (0.0001) was spent at the interim analysis because the study would not stop due to the interim results, and the sole purpose of the planned interim analyses was to adjust the sample size by using the conditional power approach under certain conditions that do not inflate the type I error (Mehta and Pocock 2011).

The sample size re-estimation process was based on the 2-sided, 2-group chi-square tests, allowing for the control of the type I error to remain at 0.05 and using a conditional power approach to retain an 85% unconditional overall study power (using methods described by Mehta and Pocock 2011). The sample size was allowed to exceed 456, which is 100% more than the planned sample size. The external independent statisticians and programmers used pre-specified decision rules, and based on observed conditional power at the interim analyses. Refer to Interim Analyses SAP for further details.

The interim analysis was performed with the data collected through 29-Nov-2017 according to the interim analysis SAP, Version 2.0 dated on 21-Nov-2017. The recommendation based on interim analysis results was that the sample size was to remain as planned without an increase. No unblinded data or results were disclosed to Shire, the PPD blinded study team, study sites, or subjects.

15. DATA MONITORING/REVIEW COMMITTEE

Not applicable

16. COMPUTER METHODS

All statistical analyses will be performed using SAS® Version 9.2 or higher (SAS Institute, Cary, NC 27513) on a suitably qualified environment

17. CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

In the Memorandum of SHP621-301 Protocol Amendment 2 – Administrative Change (dated 20 February 2019), the following changes related to statistical analysis were made to SHP621-301 Protocol Amendment 2 (dated 26 January 2018).

Protocol Section	Protocol Amendment 2 Text	Clarification
9.7 Study Population	The **intent-to-treat (ITT)** set will include all randomized subjects. Subjects will be analyzed according to their assigned treatment, regardless of the treatment actually received. The **full analysis set (FAS)** will include all randomized subjects who received at least 1 dose of a double-blind investigational product and have both an evaluable post-baseline biopsy in the treatment period (ie, peak eosinophil count is reported for at least 2 esophageal levels) and a post-baseline DSQ score.	~~The **intent-to-treat (ITT)** set will include all randomized subjects. Subjects will be analyzed according to their assigned treatment, regardless of the treatment actually received.~~ The **full analysis set (FAS)** will include all randomized subjects who received at least 1 dose of a double-blind investigational product ~~and have both an evaluable post-baseline biopsy in the treatment period (ie, peak eosinophil count is reported for at least 2 esophageal levels) and a post-baseline DSQ score~~.
9.8 Efficacy Analyses	The primary, key secondary and secondary efficacy analyses will be performed on the ITT set and presented by treatment group. Data collected at the baseline visit (Visit 1) will be used as the baseline for all efficacy analyses.	The primary, key secondary and secondary efficacy analyses will be performed on the ~~ITT~~ FAS set and presented by treatment group. Data collected at the baseline visit (Visit 1) or the last observation prior to the first dose of investigational product during double-blind treatment phase, whichever is later, will be used as the baseline for all efficacy analyses.
9.8.1 Primary Efficacy Endpoints	The co-primary efficacy endpoints will be analyzed based on the ITT set. Analyses will be repeated using the FAS and the PP set.	The co-primary efficacy endpoints will be analyzed based on the ~~ITT~~ FAS set. Analyses will be repeated using the ~~FAS and the~~ PP set. Analysis will be repeated based on FAS using different missing data handling methods to assess the impact of missing DSQ daily diary data on dysphagia symptom response.
9.8.1.1 Missing Data Imputation	The MI procedure will generate 10 version datasets with binary outcome imputed from the subjects with complete data. Once the missing values are imputed and each dataset is created, the results will be appropriately pooled across the multiply imputed estimated regression coefficients and their standard	The MI procedure will generate ~~10~~ 50 version datasets with binary outcome imputed from the subjects with complete data. Once the missing values are imputed and each dataset is created, the results will be appropriately pooled across the multiply imputed ~~estimated~~

Protocol Section	Protocol Amendment 2 Text	Clarification
	errors using PROC MIANALYZE.	~~regression coefficients~~ estimates and their standard errors using PROC MIANALYZE.
9.9 Safety Analyses	TEAEs are defined as AEs that start or deteriorate on or after the first dose of investigational product (Visit 1) and no later than 3 days following the last dose of investigational product.	TEAEs are defined as AEs that start or deteriorate on or after the first dose of investigational product (Visit 1) and ~~no later than 3 days following~~ through the Safety Follow-up Contact, or 31 days after the last dose of investigational product for subjects who do not have a Safety Follow-up Contact.

18. DATA HANDLING CONVENTIONS

18.1 General Data Reporting Conventions

Continuous variables will be summarized using the following descriptive statistics: n, mean, SD, median, Q1, Q3, minimum, and maximum. Categorical and count variables will be summarized by the number of subjects (n) and the percent of subjects in each category.

Unless specified otherwise, min/max will be presented to the same decimal places as the raw data. Percentage, mean and median will be presented to 1 more decimal places than the raw data. Standard deviation and standard error will be presented to 2 more decimal places than the raw data.

BMI should be rounded to 1 decimal place for reporting.

Derived questionnaire scores, and other similar efficacy parameters recorded as integers, should be rounded to 1 decimal place for reporting.

Averaged lab and vital sign results e.g. diastolic/systolic blood pressure and pulse (when taken in triplicate) should be rounded to 1 decimal place for reporting.

In addition, p-values will generally be presented to 3 decimal places; values less than 0.001 will be presented as <0.001.

18.2 Derived Efficay Endpoints

Refer to Section 10 for derivation instructions of histology response and DSQ scores by visit. DSQ scores including DSQ combined score and other DSQ scores will be calculated at study visits (Week 0, Week 4, Week 8, Week 12 and Week 16) using the daily diary data in the selected 14-day period prior to the actual clinical visit date. For subjects who prematurely discontinue during the Double-blind Treatment Phase, the available daily DSQ diary data collected after the actual last visit and before early termination will be included in the calculation of the diary data for the 14-day period for the next visit that was planned after ET. The target date of the next planned visit will be determined by the number of days since randomization, the next planned visit date = the randomized date + 7*[the next planned visit week number-4] + 1 (eg, for a subject who discontinued at week 15, the next planned visit week would be Week 16, and the calculated planned date for Week 16 would be determined as: [1+12 weeks*7] days after the date of randomization [Week 4 visit]. In this example, the subject who discontinued at Week 15 would have diary data included in the calculation of Week 16 DSQ scores that was within 14 days of the calculated planned date of Week 16.).

18.3 Derived Safety Endpoints

For safety parameters, the early termination visits will be mapped to the next scheduled visits. The last visits per subject will be summarized in “End of Study” visit, including both last visits of completers and early termination visits. Repeated or Unscheduled Assessments of Safety Parameters will be handled as follows:

If a subject has repeated assessments before the start of investigational product in the double-blind treatment period, then the results from the final assessment made prior to the start of investigational product will be used as baseline. If end of study assessments are repeated or unscheduled, the last post-baseline assessment will be used as the end of study assessment for generating descriptive statistics. If a subject has repeated assessments between the start of investigational product in the double-blind treatment period and the end of study visit, the assessments of unscheduled visits will be excluded in the table summary, but all the records will be included in the listings.

18.4 Missing Date of Investigational Product

When the date of the last dose of investigational product is missing for a subject in the Saety Set, all efforts should be made to obtain the date from the investigator. If it is still missing after all efforts, then the last visit date when investigational product was returned will be used in the calculation of treatment duration.

18.5 Missing Date Information for Prior or Concomitant Medications

For prior or concomitant medications, including rescue medications, incomplete (i.e., partially missing) start date and/or stop date will be imputed. When the start date and the stop date are both incomplete for a subject, impute the start date first.

18.5.1 Incomplete Start Date

The following rules will be applied to impute the missing numerical fields. If the stop date is complete and the imputed start date is after the stop date, then the start date will be imputed using the stop date.

Missing day and month

- If the year of the incomplete start date is the same as the year of the date of the first dose of investigational product, then the day and month of the date of the first dose of investigational product will be assigned to the missing fields
- If the year of the incomplete start date is before the year of the date of the first dose of investigational product, then December 31 will be assigned to the missing fields
- If the year of the incomplete start date is after the year of the date of the first dose of investigational product, then 01 January will be assigned to the missing fields.

Missing month only

- The day will be treated as missing and both month and day will be replaced according to the above procedure.

Missing day only

- If the month and year of the incomplete start date are the same as the month and year of the date of the first dose of investigational product, then the day of the date of the first dose of investigational product will be assigned to the missing day

- If either the year is before the year of the date of the first dose of investigational product or if both years are the same but the month is before the month of the date of the first dose of investigational product, then the last day of the month will be assigned to the missing day
- If either the year is after the year of the date of the first dose of investigational product or if both years are the same but the month is after the month of the date of the first dose of investigational product, then the first day of the month will be assigned to the missing day.

18.5.2 Incomplete Stop Date

The following rules will be applied to impute the missing numerical fields. If the date of the last dose of investigational product is missing, then replace it with the last visit date. If the imputed stop date is before the start date (imputed or non-imputed start date), then the imputed stop date will be equal to the start date.

Missing day and month

- If the year of the incomplete stop date is the same as the year as of the date of the last dose of investigational product, then the day and month of the date of the last dose of investigational product will be assigned to the missing fields
- If the year of the incomplete stop date is before the year of the date of the last dose of investigational product, then 31 December will be assigned to the missing fields
- If the year of the incomplete stop date is after the year of the date of the last dose of investigational product, then 01 January will be assigned to the missing fields.

Missing month only

- The day will be treated as missing and both month and day will be replaced according to the above procedure.

Missing day only

- If the month and year of the incomplete stop date are the same as the month and year of the date of the last dose of investigational product, then the day of the date of the last dose of investigational product will be assigned to the missing day
- If either the year is before the year of the date of the last dose of investigational product or if both years are the same but the month is before the month of the date of the last dose of investigational product, then the last day of the month will be assigned to the missing day
- If either the year is after the year of the last dose of investigational product or if both years are the same but the month is after the month of the date of the last dose of investigational product, then the first day of the month will be assigned to the missing day.

18.6 Missing Date Information for Adverse Events

For AEs, the default is to only impute incomplete (i.e., partially missing) start dates. Incomplete stop dates may also be imputed when calculation of the duration of an AE is required per the protocol. If imputation of an incomplete stop date is required, and both the start date and the stop date are incomplete for a subject, then the start date will be imputed first.

18.6.1 Incomplete Start Date

Rulls will be followed as described in in Section 18.5.1.

18.6.2 Incomplete Stop Date

When required per the protocol, follow the same rules as in Section 18.5.2.

18.7 Missing Severity Assessment for Adverse Events

If severity is missing for an AE starting prior to the date of the first dose of investigational product, then a severity of "Mild" will be assigned. If the severity is missing for an AE starting on or after the date of the first dose of investigational product, then a severity of "Severe" will be assigned. The imputed values for severity assessment will be used for incidence summaries, while the actual values will be used in data listings.

18.8 Missing Relationship to Investigation Product for Adverse Events

If the relationship to investigational product is missing for an AE starting on or after the date of the first dose of investigational product, a causality of "Related" will be assigned. The imputed values for relationship to double-blind investigational product will be used for incidence summaries, while the actual values will be presented in data listings.

18.9 Character Values of Clinical Laboratory Variables

If the reported value of a clinical laboratory variable cannot be used in a statistical analysis due to, for example, that a character string is reported for a numerical variable. The appropriately determined coded value will be used in the statistical analysis. However, the actual values as reported in the database will be presented in data listings.

Table 4 Examples for Coding of Special Character Values for Clinical Laboratory Variables

Clinical Laboratory Test	Possible Results (in SI units)	Possible Results (in conventional units)	Coded Value for Analysis
Chemistry: ALT	<5 U/L	<5 U/L	0
Chemistry: AST	<5 U/L	<5 U/L	0
Chemistry: Total Bilirubin	<2 umol/L	<0.15 mg/dL	0
Urinalysis: Glucose	≥50 mg/dL	≥50 mg/dL	Positive
	≤0 mg/dL	≤0 mg/dL	Negative
Urinalysis: Ketones	>0 mg/dL	>0 mg/dL	Positive
Urinalysis: Protein	>0 mg/dL	>0 mg/dL	Positive
Urinalysis: pH	≥9.0	≥9.0	9.0

19. REFERENCES

Brooks R. 1996. EuroQol: the current state of play. Health Policy. 37(1):53-72.

EuroQol 1990. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy. 16, 199-208.

Franciosi JP, Hommel KA, Bendo CB. 2013. PedsQL eosinophilic esophagitis module: feasibility,reliability, and validity. J Pediatr Gastroenterol Nutr. 57(1):57-66.

Mehta, C.R., Pocock S.J.. 2011. Adaptive increase in sample size when interim results are promising: A practical guide with examples. Statistics in Medicine. 30:3267-3284.

Taft TH,, Kern E., Kwiatek MA., Hirano I., Gonsalves N., Keefer L. 2011. The adult eosinophilic oesophagitis quality of life questionnaire: a new measure of health-related quality of life. Taft et al, Aliment Pharmacol Ther. 34: 790–798.

Wille, N., Badia, X., Bonsel, G., Burstrom, K., Cavrini, G., Devlin, N., Egmar, A. C., Greiner, W., Gusi, N., Herdman, M., Jelsma, J., Kind, P., Scalone, L. & Ravens-Sieberer, U. 2010. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Qual Life Res. 19, 875-86.

20. APPENDICES

APPENDIX 1 The Daily Dysphagia Symptom Questionnaire

1. Since you woke up this morning, did you eat solid food? Possible responses=yes , no .

2. Since you woke up this morning, has food gone down slowly or been stuck in your throat? Possible responses=yes (2), no (0).

3. For the most difficult time you had while swallowing food today, did you have to do anything to make the food go down or to get relief? Possible responses:

- No, it got better or cleared up on its own (0)
- Yes, I had to drink liquid to get relief (1)
- Yes, I had to cough and/or gag to get relief (2)
- Yes, I had to vomit to get relief (3)
- Yes, I had to seek medical attention to get relief (4)

4. The following question concerns the amount of pain you have experienced when swallowing food. What was the worst pain you had while swallowing food today? Possible responses:

- None, I had no pain (0)
- Mild (1)
- Moderate (2)
- Severe (3)
- Very Severe (4)

DSQ combined score = (Sum of points from questions 2+3 in the daily DSQ)×14 days/(Number of diaries reported with nonmissing* data).

DSQ + pain score = (Sum of points from questions 2+3+4 in the daily DSQ)×14 days/(Number of diaries reported with nonmissing* data).

DSQ pain score = (Sum of points from questions 4 in the daily DSQ)×14 days/(Number of diaries reported with nonmissing* data).

DSQ question 2 score=(Sum of points from questions 2 in the daily DSQ)×14 days/(Number of diaries reported with nonmissing* data).

DSQ question 3 score=(Sum of points from questions 3 in the daily DSQ)×14 days/(Number of diaries reported with nonmissing* data).

If answer for Question 1 is “Yes” and answer for Question 2 is “No”, then scores for Question 3 and Question 4 are set to each 0 before calculating the above scores for any analysis method proposed in the efficacy analyses.

**Data for a given day will be considered missing if the diary is not reported or if the diary is reported but the answer to Question 1 is "No". The number of diaries reported with non-missing data is also referred to as "reported diaries" in the following discussion.*

The DSQ combined score will be calculated summing the points from Questions 2 and 3 from each reported daily diary with non-missing data, and dividing this by the number of reported daily diaries with non-missing data in the selected 14-day period. This quotient will then be multiplied by 14. If subjects respond "No" to Question 2 the DSQ combined score will be set to zero for that day.

The DSQ combined score calculated during the 14-day period prior to Baseline Visit 1 will be the Baseline DSQ Score. The DSQ combined score calculated during the 14-day period prior to the Final Treatment Period Evaluation will be the Final Treatment DSQ combined score.

For each 14-day period, at least 8 reported diaries are needed to calculate a DSQ combined score. If fewer than 8 reported diaries are available within the 14-day period, then the most recent 8 reported diaries in a consecutive 14-day period would be used to calculate the DSQ combined score. In order to determine the most recent 14-day period with the minimum number of reported diaries, it may be necessary to make adjustments in some cases by shifting to earlier diary entries. Such 14-day periods cannot be shifted for more than 7 days (not including the study visit day) when calculating DSQ combined score at Baseline Visit 1 (Week 4), Visit 2 (Week 8) and Visit 3 (Week 12), and cannot be shifted for more than 14 days prior to the Visit 4 (Week 16) (not including the visit day of Visit 4) for the DSQ combined score at the Final Treatment Period Evaluation.

APPENDIX 2 Calculation of Adult Eosinophilic Esophagitis Quality of Life (EoE-QOL-A) Total Score

Adult Eosinophilic Esophagitis Quality of Life Questionnaire Standard Version (30 items)

The responses will be changed to a numeric equivalent using the following transformation: Not at all = 4, Slightly = 3, Moderately = 2, Quite a Bit = 1, Extremely = 0.

Overall score is a sum of the responses for all 30 questions (Q1-Q30) and it will have a score range from 0-120.

Subscales will be the sum of set questions:

Eating/Diet Impact will be the sum of: Q2, Q9, Q16, Q24, Q25, Q26, Q27, Q28, Q29, Q30 and it will have a score range from 0-4.

Social Impact will be the sum of: Q14, Q17, Q19, Q22 and it will have a score range from 0-16.

Emotional Impact will be the sum of: Q1, Q5, Q6, Q7, Q11, Q13, Q21, Q23 and it will have a score range from 0-32.

Disease Anxiety will be the sum of: Q4, Q10, Q12, Q15, Q18 and it will have a score range from 0-20.

Swallowing Anxiety will be the sum of: Q3, Q8, Q20 and it will have a score range from 0-12.

To compute subscale score a weighted total sum will be used. Sum up the value of the response for each of the questions in that subscale and then divide by the total number of questions answered for that subscale.

APPENDIX 3 Calculation of Pediatric Quality of Life – EoE (PedsQL-EoE, subjects 11-17 years of age) Total Score

Pediatric Quality of Life – EoE Questionnaire

The responses will be transformed to a numeric equivalent using the following reverse scoring transformation: NEVER = 100, ALMOST NEVER = 75, SOMETIMES = 50, OFTEN = 25, ALMOST ALWAYS = 0.

We will create dimensions by combining different questions to form the dimensions (Symptoms I, Symptoms II, Treatment, Worry, Communication, Food and Eating, Food Feelings, and Feeding Tube). Feeding Tube dimension will not be used for reporting or analyses.

SYMPTOMS I will be the sum of CRF questions ("Burning in chest", "Chest pain, ache, or hurt", "Feel like Throwing up", "Food coming back up throat", "Stomach aches or belly aches", "Throwing up").

SYMPTOMS II will be the sum of CRF questions ("Need drink to help swallow food", "Needing more time to eat", "Trouble swallowing", "food stuck in throat or chest").

TREATMENT will be the sum of CRF questions ("Getting allergy testing", "Getting an endoscopy", "Going to the doctor", "Not wanting to take medicines", "Remembering to take medicines").

WORRY will be the sum of CRF questions ("Getting allergy testing", "Getting an endoscopy Worry", "Getting sick in front of other people", "Going to the doctor", "Having EoE", "other people think about me because of EoE").

COMMUNICATION will be the sum of CRF questions ("Tell people about EoE", "Telling adults how feels", "Telling nurses how feels", "Telling friends how feels", "Telling parents how feels").

FOOD AND EATING will be the sum of CRF questions ("following diet restriction", "not eating same as family", "not eating same as friends", "sneaking food")

FOOD FEELINGS will be the sum of CRF questions ("Feeling Mad", "Feeling Sad", "Worry about foods allergic too").

FEEDING TUBE will be the sum of CRF questions ("Remember to use feeding tube", "Using a feeding tube").

The dimension score will not be calculated if more than 50% of the items in the scale are missing. To compute dimension score a weighted total sum will be used. Sum up the value of the response for each of the questions in that dimension and then divide by the total number of questions answered for that dimension.

A total score will be the sum of all the items scores except for FEEDING TUBE.

Symptoms total scale score will be calculated as the sum of items scores when dimensions in Symptoms I and II divided by the number of items answered.

APPENDIX 4 Calculation of the EQ-5D Score

ED-5D 3L and ED-5D Y Composite Index Score for US-Based Population

The US population-based EQ-5D index score ranges from -0.11 to 1.0 on a scale where 0.0 = death and 1.0 = perfect health. The negative index scores are regarded as worse than death. The calculations of the EQ-5D index scores are based on a regression equation derived from a large-scale survey of the general adult US population (Shaw 2005). The regression equation for scoring is as follows:

X = 1 - 0.146 MO2 - 0.558 MO3 - 0.175 SC2 - 0.471 SC3 - 0.140 UA2 - 0.374 UA3 - 0.173 PD2 - 0.537 PD3 - 0.156 AD2 - 0.450 AD3 + 0.140 D1 - 0.011 I2-squared + 0.122 I3 + 0.015 I3-squared

Where X is the US EQ-5D index score and the independent variables are listed below

MO2: 1 if mobility is level 2; 0 otherwise

MO3: 1 if mobility is level 3; 0 otherwise

SC2: 1 if self-care is level 2; 0 otherwise

SC3: 1 if self-care is level 3; 0 otherwise

UA2: 1 if usual activities is level 2; 0 otherwise

UA3: 1 if usual activities is level 3; 0 otherwise

PD2: 1 if pain/discomfort is level 2; 0 otherwise

PD3: 1 if pain/discomfort is level 3; 0 otherwise

AD2: 1 if anxiety/depression is level 2; 0 otherwise

AD3: 1 if anxiety/depression is level 3; 0 otherwise

D1: number of dimensions beyond the first in level 2 or level 3 (i.e. 0 if 0, 1 if 1 dimension in level 2 or 3 or number of dimensions in level 2 or level 3 minus 1 otherwise)

I2-squared: squared number of dimensions beyond the first in level 2

I3: number of dimensions beyond the first in level 3

I3-squared: squared I3

APPENDIX 5 Histopathologic Epithelial Features Combined Total Score (Grade and Stage)

At each of the esophageal levels (proximal, mid and distal), histopathologic features are scored for both Grade and Stage. The histopathologic epithelial features consist of eosinophil peak, basal layer hyperplasia, eosinophil abcesses. surface layering, dilated intercellular spaces, surface alteration, dyskeratotic epithelial cells, and lamina propria fibrosis. Each of the 8 Histopathologic features at each eosophageal level has a possible score 0-3 points for both Grade (severity of abnormal histologic feature) and Stage (extend of abnormal histologic feature). Combined total scores include the following parameters for grade and stage respectively: 1, combined total score ratio (TSR) =(proximal TSR + mid TSR + distal TSR)/N, where N is the number of non missing sections for TSR; 2, combined score for each individual histopathologic feature (proximal+mid+distal)/N, where N is the number of non-missing sections for each individual feature.