STATISTICAL ANALYSIS PLAN

VERSION: 1.0

DATE OF PLAN: 14 January 2021

STUDY DRUG:

IW-3718

PROTOCOL NUMBER:

C3718-301

STUDY TITLE:

A Phase 3, Randomized, Double-blind, Placebo-controlled, Parallel-group, Multicenter Trial of Oral IW-3718 Administered to Patients with Gastroesophageal Reflux Disease while receiving Proton Pump Inhibitors

SPONSOR:

Ironwood Pharmaceuticals, Inc.

100 Summer Street, Suite 2300

Boston, MA 02110

[redacted]

This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Term
AE	adverse event
ANCOVA	analysis of covariance
BID	twice daily
BP	blood pressure
BPM	beats per minute
CDF	cumulative distribution function
CI	confidence interval
CMH	Cochran Mantel-Haenszel
CRO	contract research organization
DHSS	Daily Heartburn Severity Score
DPP	Data Presentation Plan
DRFS	Daily Regurgitation Frequency Score
eCRF	electronic case report form
eDiary	electronic diary
ECG	electrocardiogram
EE	erosive esophagitis
EGD	esophagogastroduodenoscopy
EOT	end of treatment
EQ	EuroQol
GERD	gastroesophageal reflux disease
HEENT	head, ears, eyes, nose, throat
ICF	informed consent form
ICH	International Conference on Harmonisation
IPD	important protocol deviations
IWRS	interactive web response system
LLN	lower limit of normal
IRT	interactive response technology
LS	least squares
MAR	missing at random

MCMC	Monte-Carlo Markov Chain
MedDRA	Medical Dictionary for Regulatory Activities Terminology
MI	multiple imputation
mITT	modified intent-to-treat
MMRM	mixed model for repeated measures
MNAR	missing not at random
mRESQ-eD	modified Reflux Symptom Questionnaire - electronic Diary
non-EE	without erosive esophagitis
OHR	Overall Heartburn responder
PCS	potentially clinically significant
PHFD	proportion of heartburn-free days
PT	preferred term
PID	patient identification number
PPI	proton pump inhibitors
QD	once daily
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SI	Le Système International d'Unités (International System of Units)
SOC	system organ class
TEAE	treatment-emergent adverse event
ULN	upper limit of normal
US	United States
WHSS	Weekly Heartburn Severity Score
WPAIQ	Work Productivity and Activity Impairment Questionnaire
WRFS	Weekly Regurgitation Frequency Score

1. INTRODUCTION

Study C3718-301 is a Phase 3, randomized, double-blind, placebo-controlled, parallel-group, multicenter study of oral IW-3718 administered to patients with gastroesophageal reflux disease (GERD) while receiving proton pump inhibitors (PPIs).

This statistical analysis plan (SAP) describes the statistical analysis methods for all data collected according to the following documents:

- Study Protocol C3718-301 (Amendment 2, dated 14 August 2020).
- Electronic Case Report Form (eCRF) Version 1.0 (dated 27 May 2020).
- Electronic Dairy (eDiary) (dated 10 July 2018)

Specifications for tables, figures, and listings are contained in a separate standard document, the Data Presentation Plan (DPP).

2. STUDY OBJECTIVES

The objective of this study is to evaluate the safety and efficacy of IW-3718 administered to patients with gastroesophageal reflux disease (GERD) who continue to have persistent symptoms, such as heartburn and regurgitation, while receiving once-daily (QD), standard-dose proton pump inhibitors (PPIs).

3. STUDY DESIGN

3.1. General Description

Study C3718-301 is a multicenter, randomized, double-blind, placebo-controlled, parallel-group, 8-week study, consisting of 4 distinct periods as illustrated in Figure 1. The study will enroll patients who have GERD and continue to experience GERD symptoms while receiving QD, standard-dose PPI therapy. Eligible patients will continue to take their PPI and will be randomized in equal proportions (1:1) to 1500-mg IW-3718 or matching placebo twice daily (BID).

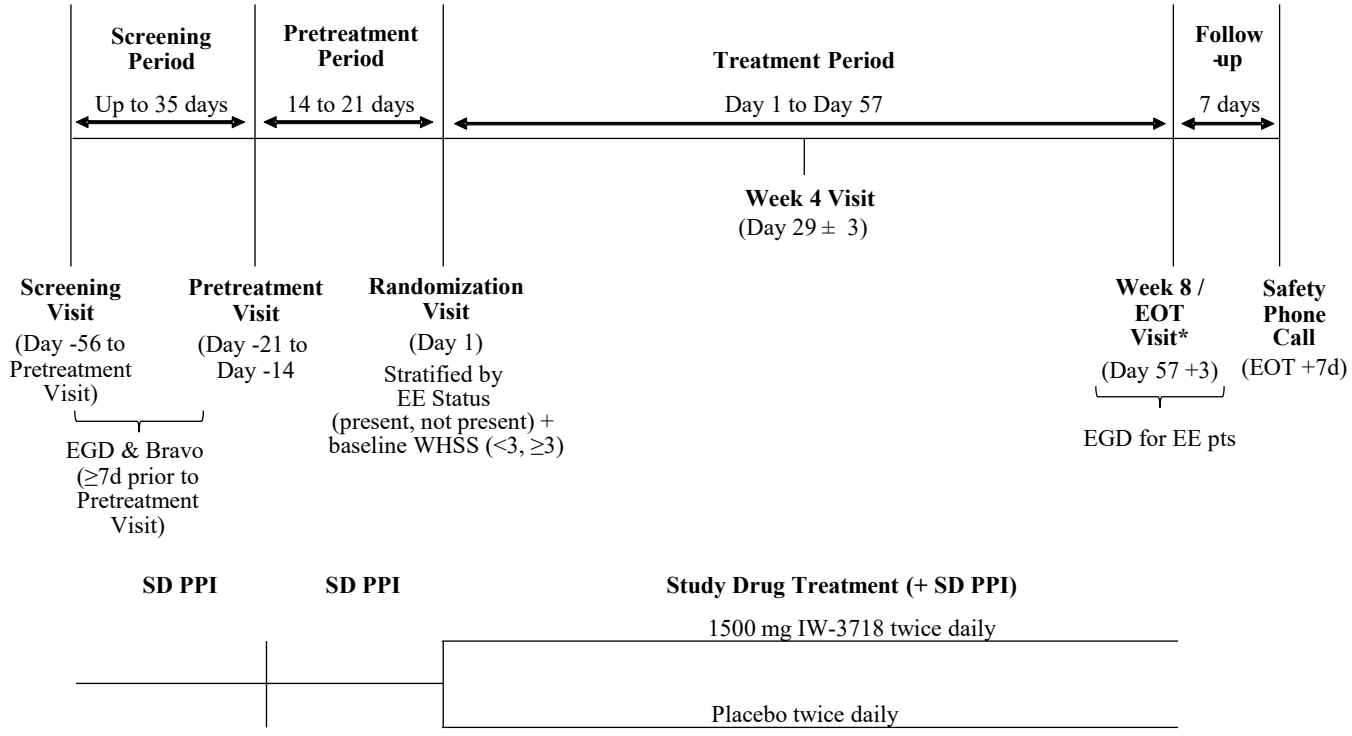
The total duration of patient participation is expected to last up to 18 weeks, including up to 35 days of Screening, 14 to 21 days of the Pretreatment Period, 8 weeks of the double-blind Treatment Period, and 7 days of Follow-up. During the Screening Period, patients undergo an esophagogastroduodenoscopy (EGD) and up to 96 hours of pH testing with the Bravo™ device while taking their standard-dose PPI. During the Pretreatment Period, patients will continue to take their PPI and will provide the daily and weekly symptom assessments using a handheld electronic diary (eDiary). Patients meeting all the entry criteria for this trial at the end of the Pretreatment Period will enter the double-blind Treatment Period. During the Treatment Period, patients 1) will continue to take their usual label-dose PPI approximately 30-60 minutes before breakfast each day, 2) will take the double-blind study drug (IW-3718 or matching placebo) immediately after the morning and evening meals each day, and 3) will use the eDiary to provide their daily and weekly symptom assessments. Patients who had erosive esophagitis (EE) at Screening and completed ≥4 weeks of treatment will have a repeat EGD at the End of Treatment (EOT) Visit. The study site will contact all patients via telephone 7 days following the EOT Visit to collect information regarding adverse events (AEs) and concomitant medications/procedures.

3.2. Discussion of Study Design, Including Choice of Control Group

A randomized, double-blind, placebo-controlled, parallel-group, multicenter study design was chosen in accordance with the concepts in International Conference on Harmonisation (ICH) E10, Choice of Control Groups and Related Issues in Clinical Trials (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2001), in order to avoid baseline characteristics selection bias for different treatment groups on the part of the participants and investigators. Blinded treatment is used to reduce potential bias during data collection and evaluation of clinical results. The placebo arm is used to control the subject's expectation that a treatment would have an effect, to optimize sensitivity, to maintain the scientific rigor of the study, and to validate the study internally.

This study has a 14- to 21-day Pretreatment Period to establish a baseline without test therapy and to familiarize patients with data collection methodology (ie, eDiary), and an 8-week Treatment Period to compare the test treatment with a placebo control. Finally, there is a 7-day follow-up phone call.

Figure 1: Overview of Study Design



Note: There is no Day 0

* This visit represents the end of the study
EE = erosive esophagitis; EGD=esophagogastroduodenoscopy; EOT=end of treatment; SD PPI = standard dose proton pump inhibitor; WHSS = Weekly Heartburn Severity Score

3.3. Treatment Administered

During the Pretreatment and Treatment Periods, all patients will take their current prescribed, label-dose PPI. The PPI will be taken each day, approximately 30-60 minutes before breakfast.

During the Treatment Period, patients will be instructed to take study drug (IW-3718 or matching placebo) in the morning (immediately upon completion of breakfast) and in the evening (immediately upon completion of dinner) each day, except for the first dose, which will be taken with liquid and a snack in clinic at the Randomization Visit on Day 1.

Lastly, during the Pretreatment and Treatment Periods, patients may also take site-dispensed, protocol-permitted antacid as rescue medicine when their heartburn becomes intolerable.

3.4. Assigning Patient to Treatment Group

Patients who meet all the inclusion criteria and none of the exclusion criteria will be randomized into the study at the Randomization Visit on Day 1. Approximately 550 patients will be stratified by 1) whether they have or do not have EE on the screening EGD and 2) their baseline heartburn severity level (<3 vs. ≥3); and randomized through central randomization in a 1:1 ratio to either 1500-mg IW-3718 BID (while taking their usual label-dose PPI) or placebo BID (while taking their usual label-dose PPI).

The computer-generated randomization schedule was prepared by an independent statistician not otherwise associated with the study. Randomization numbers will be assigned by interactive web response system (IWRS).

3.5. Blinding

This is a double-blind study. Ironwood Pharmaceuticals, Inc. (Ironwood, the Sponsor) study personnel, the site investigators, all other site study personnel, and the patient will remain blinded to individual patient treatment assignments throughout the study. Procedures for unblinding in the case of an emergency are included in Clinical Study Protocol C3718-301, Section 3.4.7.

4. DETERMINATION OF SAMPLE SIZE

The sample size of 660 patients (330 patients per treatment group) was chosen to ensure adequate power for using the fixed sequence testing method for the primary and key secondary efficacy parameters.

The power calculations for the primary and key secondary efficacy parameters are based on results from patients with evidence of pathological acid reflux (positive baseline Bravo status) in Study ICP-3718-202 that had a similar study population as the current study.

The primary efficacy parameter of the study is the Change from Baseline at Week 8 in Weekly Heartburn Severity Score (WHSS), and key secondary efficacy parameters include:

- Change from Baseline at Week 8 in Weekly Regurgitation Frequency Score (WRFS)
- Overall Heartburn Responder (OHR)
- Proportion of Heartburn-free Days during the 8-week Treatment Period (PHFD)

For definition of the efficacy parameters above, refer to Section 7.9.1.2 and Section 7.9.2.

Using data from patients in Study ICP-3718-202 with similar inclusion/exclusion criteria as the current study, ie, subjects with a positive baseline BRAVO status, we derived the estimated results at end of the study as follows:

Table 1: Study ICP-3718-202 - Summary of Efficacy Parameters at Week 8 (BRAVO+)

Order of Hypothesis Testing	Parameters	Placebo BID (+ PPIs)	1500 mg IW-3718 BID (+ PPIs)
1	Change from Baseline at Week 8 in WHSS	N = 43 Mean = -1.58 SD = 1.20	N =51 Mean = -2.01 SD = 1.29
2	Change from Baseline at Week 8 in WRFS	N = 43 Mean = -0.81 SD = 1.12	N = 51 Mean = -1.26 SD = 0.91
3	Overall Heartburn Responder	18/49 = 36.7%	31/55 = 56.4%
4	Proportion of Heartburn-free Days during the 8-week Treatment Period	N = 48 Mean = 0.195 SD = 0.2822	N = 55 Mean = 0.268 SD = 0.2928

Based on a resampling with replacement-based bootstrap simulations, controlling for multiplicity, assuming there are approximately 45% patients with erosive esophagitis (EE), a sample size of 330 patients per group will provide approximately 99% power to detect a treatment difference between 1500-mg IW-3718 (while receiving PPIs) and placebo (while receiving PPIs) for the primary efficacy parameter, assuming the randomized population in this study is consistent with the selected population from Study ICP-3718-202. Power for other key efficacy parameters is listed in Table 2 below. The 2-sample t-test was used for continuous parameters and the Chi-square test was used for the responder parameter.

Table 2: Hypothesis Testing Order and Power for Fixed Sequence Testing

Order of Hypothesis Testing	Parameters	Power for Fixed Sequence Testing
1	Change from Baseline at Week 8 in WHSS	99%
2	Change from Baseline at Week 8 in WRFS	99%
3	Overall Heartburn Responder	99%
4	Proportion of Heartburn-free Days during the 8-week Treatment Period	88%

These efficacy parameters will be tested sequentially according to the above order at a 2-sided significance level of 5%, so that the overall type I error rate is maintained at 5%.

5. PHARMACOKINETICS, EFFICACY, AND SAFETY ASSESSMENTS

5.1. Pharmacokinetic Assessments

There are no pharmacokinetic assessments planned for this study.

5.2. Efficacy Assessments

The efficacy assessment instruments that will be used to derive the efficacy parameters are listed along with the instrument names, questionnaires, measurement scales, references, and associated efficacy parameter(s) in Table 3.

Table 3: Efficacy Assessment Instruments and Parameters

Instruments	Parameters	Questionnaires	Measurement Scales	Reference
mRESQ-eD[a]	• Weekly Heartburn Severity Score (WHSS) • Overall Heartburn Responder (OHR)	*Item 1* - Burning feeling behind the breastbone or in the center of the upper stomach. *Item 2* - Pain behind the breastbone or in the center of the upper stomach	0 = Did not have, 1 = Very mild, 2 = Mild, 3=Moderate, 4 = Moderately severe, 5 = Severe.	Appendix 5
	Weekly Regurgitation Frequency Score (WRFS)	*Item 6* - Regurgitation (liquid or food moving upwards toward your throat or mouth) *Item 7* - An acid or bitter taste in the mouth	0 = Never, 1 = Rarely, 2 = Sometimes, 3 = Often, 4 = Very often	Appendix 5
	Proportion of Heartburn-free Days during the 8-week Treatment Period (PHFD)	*Item 1* - Burning feeling behind the breastbone or in the center of the upper stomach	0 = Did not have	Appendix 5
	GERD Symptoms	*Item 3* - Difficulty swallowing *Item 4* - Hoarseness *Item 5* - Cough	0 = Did not have, 1 = Very mild, 2 = Mild, 3 = Moderate, 4 = Moderately severe, 5 = Severe	Appendix 5
	GERD symptom - Burping	*Item 8* - Burping	0 = Never, 1 = Rarely, 2 = Sometimes, 3 = Often, 4 = Very often	Appendix 5
Other Daily Assessments	Sleep Disturbance Due to GERD - Falling Asleep	Last night, did you have trouble falling asleep because of GERD symptoms?	Yes, No	Appendix 6
	Sleep Disturbance Due to GERD - Awakenings	Last night, how many times did you wake up during the night because of GERD symptoms?	0, 1 time, 2 times, 3-5 times, 6 or more times	Appendix 6
	Use of rescue medicine (antacid)	How many times did you use your rescue medicine (liquid antacid) during the past 24 hours?		Appendix 6

Instruments	Parameters	Questionnaires	Measurement Scales	Reference
	Heartburn score	Over the past 24 hours, how would you rate the severity of your heartburn	0 = Did Not Have, 1 = Very Mild, 2 = Mild, 3 = Moderate, 4 = Moderately Severe, 5 = Severe	Appendix 6
Weekly Assessments	Degree of Relief	• How would you rate your heartburn (a pain or burning sensation in your chest, behind the breastbone) over the past 7 days? • How would you rate your regurgitation (the feeling of stomach contents, either liquid or food, moving upwards to your throat or mouth) over the past 7 days? • How would you rate your overall GERD symptoms over the past 7 days?	1 = Significantly relieved, 2 = Moderately relieved, 3 = Somewhat relieved, 4 = Unchanged, 5 = Somewhat worse, 6 = Moderately worse, 7 = Significantly worse	Appendix 7
	Bothersomeness	• How much were you bothered by heartburn (a pain or burning sensation in your chest, behind the breastbone) over the past 7 days? • How much were you bothered by regurgitation (the feeling of stomach contents, either liquid or food, moving upwards to your throat or mouth) over the past 7 days? • How much were you bothered by your overall GERD symptoms over the past 7 days?	1 = Not at all, 2 = A little bit, 3 = A moderate amount, 4 = A great deal, 5 = An extreme amount	Appendix 7

Instruments	Parameters	Questionnaires	Measurement Scales	Reference
	Treatment Satisfaction	How would you rate your satisfaction with the study treatment?	1 = Very dissatisfied, 2 = Dissatisfied, 3 = Neither satisfied nor dissatisfied, 4 = Satisfied, 5 = Very satisfied	Appendix 7
Los Angeles classification of esophagitis	• Erosive esophagitis healed • Erosive esophagitis improved by at least one grade		1. Not present = No breaks (erosions) in the esophageal mucosa. (However, edema, erythema, or friability may be present.) 2. Grade A = One (or more) mucosal break no longer than 5 mm that does not extend between the tops of two mucosal folds. 3. Grade B = One (or more) mucosal break more than 5 mm long that does not extend between the tops of two mucosal folds. 4. Grade C = One (or more) mucosal break that is continuous between the tops of two or more mucosal folds, but which involve less than 75% of the circumference. 5. Grade D = One (or more) mucosal break which involves at least 75% of the esophageal circumference.	Appendix 8

a. mRESQ-eD = modified Reflux Symptom Questionnaire – electronic Diary.

5.2.1. Primary Efficacy Assessment

The daily patient assessments of heartburn, using Items 1 and 2 of the mRESQ-eD (Table 3) and recorded once daily in the evening before bed via eDiary, will be used to derive the primary efficacy parameter WHSS (see Section 7.9.1.2).

Baseline WHSS is derived based on the patient's daily eDiary responses to the mRESQ-eD questionnaire, using the web-based system provided by CRF Health. There are 2 possible derivations for the baseline WHSS:

1. CRF Health-derived WHSS
2. Sponsor-derived WHSS

The Sponsor-derived WHSS baseline (Item 2 in the list above) will be used for all analyses.

5.2.2. Secondary Efficacy Assessments

The daily patient assessments of regurgitation, using Items 6 and 7 of the mRESQ-eD (Table 3) and recorded once daily in the evening before bed via eDiary, will be used to determine secondary efficacy parameter WRFS. Items 1 and 2 of the mRESQ-eD (Table 3) will be used to determine key secondary efficacy parameter OHR, and Item 1 of the mRESQ-eD (Table 3) will be used to derive the key secondary efficacy parameter PHFD.

5.2.3. Other Efficacy Assessments

The efficacy assessments of Items 3, 4, 5 and 8 of the mRESQ-eD (Table 3), together with the assessments below, are used in determining the additional efficacy parameters.

Other Daily Assessments

1. Additional GERD symptom assessments are completed by patients once daily in the evening before bed via eDiary (Table 3):
   - Difficulty swallowing (Item 3 of the mRESQ-eD)
   - Hoarseness (Item 4 of the mRESQ-eD)
   - Cough (Item 5 of the mRESQ-eD)
   - Burping (Item 8 of the mRESQ-eD)
2. Assessment of sleep disturbance due to GERD symptoms (Daily Sleep Disturbance due to GERD [Table 3]) are completed by patients once daily upon getting up each morning via eDiary:
   - Last night, did you have trouble falling asleep because of GERD symptoms?
   - Last night, how many times did you wake up during the night because of GERD symptoms?
3. Use of per-protocol rescue medicine (Table 3) is completed by patients once daily in the evening before bed via eDiary:
   - How many times did you use your rescue medicine (liquid antacid) during the past 24 hours?

4. Heartburn score is rated by patients once daily in the evening before bed via eDiary (Table 3)

Weekly Assessments

The following information is completed by patients once each week during the evening before bed via eDiary (Table 3):

1. Degree of Relief Assessment
2. Bothersomeness Assessment
3. Treatment Satisfaction Assessment

EGD Assessments

The following information is collected by the review the EGDs, obtained at the times specified in the protocol, according to the Los Angeles Classification of Esophagitis (Table 3):

1. Erosive Esophagitis Healing Assessment

5.3. Health Outcomes Assessments

WPAI-Sleep Disturbance-GERD

The Work Productivity and Activity Impairment Questionnaire: Sleep Disturbance-GERD (WPAI-Sleep-GERD) include questions about time lost from work, reduced productivity while at work, and reduced productivity while doing regular daily activities resulting from sleep disturbance due to GERD symptoms. Patients will complete the WPAI-Sleep-GERD at the Randomization Visit, the Week 4 Visit, and at the EOT Visit and will record their responses in the eDiary.

EQ-5D-3L

The EuroQol (EQ)-5D-3L is a generic measure of health widely used in Europe (2). The first component consists of 5 questions assessing the following dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Responses to the 5 questions define a health state. The second component of the EQ-5D is a visual analogue scale, asking patients to rate their health from 0 to 100 (0 represents worst imaginable health state and 100 represents best imaginable health state). Patients will complete the EQ-5D-3L at the Randomization Visit, the Week 4 Visit, and at the EOT Visit and will record their responses in the eDiary.

5.4. Other Assessments

Esophagogastroduodenoscopy

All patients will be required to undergo an EGD during the Screening Period. There should be a minimum of 7 days between the EGD and the start of the Pretreatment Period to allow for pH testing and patient stabilization. An EGD will be performed at the Week 8/EOT Visit for all patients who have completed at least 4 weeks of treatment and had erosive esophagitis on the screening EGD (based on the Los Angeles Classification of Esophagitis, Appendix 8) as determined by either the site personnel or the central reader.

Baseline EE status is determined using images created during a screening EGD by the local reads. These images are read and interpreted both by local physicians and centrally. There were 3 different approaches to determine the baseline EE status:

1. The site coordinator used the web-based reporting system provided by Suvoda to enter the EE information into the Interactive Response Technology (IRT) system. This information was used for randomization.
2. The site coordinator entered the local reads into the eCRF.
3. Central review: In some cases, the images received by the central reviewers were of poor quality and the central-reviewing physicians were unable to read the images. Thus, as a practical matter, the values were set to missing. The central reads are not used to overturn the local reads, and the central reads are not used anywhere in the study.

The local reads (Item 2 in the list above) of EE status will be used for statistical inference, while the central reads (Item 3 in the list above) will be used for the esophageal erosion healing analysis only.

5.5. Safety Assessments

Safety will be evaluated by AEs and serious adverse events (SAEs), standard clinical laboratory assessments, vital signs, physical examinations, medical and disease history, and electrocardiograms (ECGs). Planned timepoints for all safety assessments are provided in the Schedule of Evaluations (Appendix 4).

5.5.1. Adverse Events

An AE is any untoward medical occurrence associated with the use of a drug in humans, whether or not considered drug related.

An AE, therefore, can be any unfavorable and unintended sign (including an abnormal laboratory finding, for example), symptom, or disease temporally associated with the use of a medicinal product, whether or not considered related to the medicinal product. A procedure is not an AE, but the reason for a procedure may be an AE.

A treatment-emergent adverse event (TEAE) is an event that emerges, or a preexisting event that worsens (with a new start date), any time after administration of the first dose of study drug during the treatment emergent period, which is defined as the date and time of the first treatment to the date and time on or ≤ 1 day of the last treatment. In fact, there is an Investigator assigned flag on the eCRF to indicate an TEAE. The Investigator will provide an assessment of the AE relationship to study drug and AE severity for each AE.

The AE assessment (interview) will be collected from the time the patient signed the informed consent until completion of the Follow-up Phone Call. At each trial visit, the investigator will inquire about the occurrence of AEs/SAEs since the last visit. In addition, certain laboratory abnormalities and changes in vital signs, physical examination findings, and 12-lead ECG parameters are considered AEs per protocol if the Investigator considers them clinically significant and/or they necessitate intervention. These abnormal findings will be captured on the AE page of the electronic case report form (eCRF).

5.5.1.1. Adverse Events Causality Assessment

For all AEs, the investigator must provide an assessment of causal relationship to the study drug. The investigator must assess the relationship of each AE (including SAEs) to the use of a study drug using a 2-category scale (not related or related) based on clinical judgment and using all available information, and may include consideration of the following factors:

- Possible alternative causes of the AE, including the disease under treatment, preexisting conditions, concomitant use of other drugs, and presence of environmental or genetic factors
- The temporal association between drug exposure and onset of the AE
- Whether the manifestations of the AE are consistent with known actions or toxicity of the study drug
- Whether the AE resolved or improved with stopping use of the study drug.
- Judgment should be used if multiple products are discontinued at the same time.

The causality assessment must be recorded in the patient's source documentation and on the AE page of the patient's eCRF. The causal relationship between study drug and the AE will be assessed using the categories shown in the following table:

Category	Definition
Not related	An AE is not associated with study medication if: - Lack of a temporal relationship to study drug administration makes a causal relationship improbable (eg, the event did not occur within a reasonable time frame following administration of the study medication); and/or - Other causative factor(s) (eg, a preexisting clinical condition, other concomitant treatments) more likely explain the occurrence of the event, and - The event did not improve with stopping of the investigational product, and/or - The event did not recur upon re-exposure with investigational product
Related	An AE is attributed to the study medication if: - A temporal relationship to study drug administration makes a causal relationship plausible (eg, the event occurred within a reasonable timeframe following administration of study medication); and/or - Other causative factor(s) (eg, the patient's clinical condition, other concomitant treatments) either do not explain the event or are less equally likely to have led to the occurrence of the event, or - The event improved with stopping of the investigational product, and/or - The event recurred upon re-exposure with investigational product

5.5.1.2. Classification of Adverse Event Severity

The investigator or delegated physician will provide an assessment of the severity of each AE by recording a severity rating in the patient's source documentation and on the AE page of the patient's eCRF. *Severity*, which is a description of the intensity of manifestation of the AE, is distinct from *seriousness*, which implies a patient outcome or AE-required treatment measure

associated with a threat to life or functionality. Severity will be assessed according to the following scale by the investigators:

Mild	A type of AE that is usually transient and may require only minimal treatment or therapeutic intervention. The event does not generally interfere with usual activities of daily living.
Moderate	A type of AE that is usually alleviated with additional specific therapeutic intervention. The event interferes with usual activities of daily living, causing discomfort, but poses no significant or permanent risk of harm to the research participant.
Severe	A type of AE that interrupts usual activities of daily living, or significantly affects clinical status, or may require intensive therapeutic intervention.

AEs should be recorded using the maximum intensity of the event (eg, if a patient reported nausea lasting 3 days, 1 start date and 1 stop date should be recorded along with the maximum intensity experienced for that event over that 3-day timeframe).

5.5.1.3. Serious Adverse Events

All adverse events will be categorized as Serious or Non-Serious. An AE is considered "serious" if, in the view of either the Investigator or Sponsor, it results in any of the following outcomes:

- Death
- Life-threatening experience
- Inpatient hospitalization or prolongation of existing hospitalization
- A persistent or significant disability/incapacity or substantial disruption of the ability to conduct normal life functions
- A congenital anomaly or birth defect

Is considered an important medical event: It may not result in death, be life-threatening, or require hospitalization may be considered serious when, based upon appropriate medical judgment, they may jeopardize the subject and may require medical or surgical intervention to prevent one of the outcomes listed in this definition.

5.5.2. Medical and Disease History

A complete medical and disease history will be performed as defined in the Schedule of Evaluations (Appendix 4).

5.5.3. Physical Examination, Body Weight, and Height

A complete physical examination will be performed as defined in the Schedule of Evaluations (Appendix 4). A physical examination will include the following assessments: general appearance; head, ears, eyes, nose, and throat (HEENT); cardiovascular system; neck; respiratory system; musculoskeletal system; abdomen/liver/spleen; skin; lymph nodes; mental status; and neurologic status. Breast, genitourinary, and rectal examinations are optional at the discretion of the investigator.

Each patient's weight will be recorded at every study visit; height will only be recorded at the Screening Visit.

5.5.4. Vital Signs

Vital sign measurements will be performed as defined in the Schedule of Evaluations (Appendix 4). Respiratory rate, pulse, and blood pressure readings will be taken after the patient has been seated for at least 5 minutes.

5.5.5. Electrocardiograms

A 12-lead ECG will be performed as defined in the Schedule of Evaluations (Appendix 4). ECGs should be obtained after the patient has been supine for at least 5 minutes.

5.5.6. Clinical Laboratory Data

Blood and urine samples for clinical laboratory tests will be collected at the visits defined in the Schedule of Evaluations (Appendix 4). The clinical laboratory tests for hematology, chemistry, urinalysis, coagulation, pregnancy test and drug screening will be performed. For details refer to Clinical Study Protocol C3718-301, Section 3.5.6.

6. STATISTICAL METHODS

6.1. General Methods

In general, descriptive statistics will be presented by treatment group and by visit, as applicable. Descriptive statistics including the number of patients, mean, standard deviation (SD), median, minimum, and maximum will be calculated for continuous variables. Frequencies and percentages for each category will be calculated for categorical variables. All hypothesis tests will be 2-sided with a 5% significance level, and 95% confidence intervals (CIs) will be used.

Baseline EE status (EE vs. non-EE) that is based on local reader's assessments using images created during a screening EGD, and baseline WHSS (<3 vs. ≥3) per eDiary assessments by mRESQ-ed questionnaire are used for analysis, whenever it is appropriate.

All statistical analyses will be performed using SAS Version 9.4 (or later).

6.2. Adjusting for Baseline Characteristics and Covariates

Analyses of efficacy parameters will be adjusted by randomization stratification variables baseline EE status (EE vs. non-EE) and baseline WHSS (<3 vs. ≥3).

As a general approach, the baseline value will be included in the linear mixed effects-model for repeated measures (MMRM) and Poisson model analysis as a covariate when the postbaseline or change from baseline values are analyzed, and the randomization stratification variables baseline EE status and baseline WHSS (<3 vs. ≥3) will be adjusted in the analysis model as well. Heterogeneity of slopes will be explored, when appropriate.

6.3. Handling of Dropouts/Missing Data Imputation Method

The WHSS (Section 7.9.1.2) is the average of non-missing Daily Heartburn Severity Scores (DHSS) for the week (Appendix 1). If a patient who reported DHSS for fewer than 1 day during a week within the Treatment Period for a given week, then WHSS is missing for that week.

The WRFS (Section 7.9.2) for a week is the average of the non-missing Daily Regurgitation Frequency Scores (DRFS) for that week (Appendix 1). If a patient who reported DRFS for fewer than 1 day during a week within the Treatment Period for a week, then WRFS is missing for the week.

For responder parameters, OHR (Section 7.9.2), a patient who reported daily symptom severity scores for fewer than 4 days during a week within the Treatment Period will be considered a non-responder for that week. A patient will be deemed a weekly non-responder for all weeks following the patient's discontinuation from the study.

Unless stated otherwise, missing data will not be imputed for any of the efficacy parameters.

6.4. Interim Analysis and Data Monitoring

No interim analyses are planned to compare treatment groups with respect to efficacy or safety prior to formal completion of the trial.

6.5. Multicenter Studies

This study is being conducted in approximately 100 centers. For reporting purpose, study centers will be pooled into the following 6 geographic regions (as listed in Appendix 3): US Northeast, US Southeast, US Midwest, US Southwest, US West and Canada.

6.6. Multiple Comparisons/Multiplicity

The overall family-wise Type I error rate for the primary and key secondary efficacy analyses will be controlled at the $\alpha = 5\%$ level by employing a fixed sequence testing method. If the primary hypothesis for the comparison between the placebo group and the IW-3718 group is statistically significant ($\alpha < 5\%$) in the mITT Population (defined in Section 7.1), then the primary objective of the trial will have been achieved and the next hypothesis can be tested; otherwise, testing will stop. The testing of each subsequent hypothesis is conditioned upon all the previous hypotheses being rejected at the 5% level of significance. If a hypothesis is not rejected at the 5% level of significance, then all remaining hypotheses are deemed not statistically significant.

Table 4: Testing Order of Efficacy Parameters for the Fixed Sequence Testing

Testing Order	Parameters
1	Change from Baseline in WHSS at Visit 5 (Week 8)
2	Change from Baseline in WRFS at Visit 5 (Week 8)
3	Overall Heartburn Responders
4	Proportion of Heartburn-free Days During the 8-week Treatment Period

6.7. Examination of Subgroups

Subgroup analyses will be performed on the primary efficacy parameter, using the subgroups defined in Table 5.

Table 5: Subgroup Variables

Subgroup Variables	Subgroups
Age (years)	18 to <40; 40 to <65; ≥ 65
Sex	Female; Male
Race	Caucasian; Black/African American; Other
Ethnic Origin	Hispanic; Non-Hispanic
Body Mass Index (BMI)	<30 kg/m^2; ≥30 kg/m^2
Baseline EE Status	EE; non-EE
Baseline WHSS	<3; ≥ 3

7. ANALYSIS METHODS

7.1. Analysis Populations

- **Screened Population** consists of all patients who signed informed consent and received a patient identification (PID) number.
- **Randomized Population** consists of all patients who were assigned to a treatment group (placebo or IW-3718) via randomization.
- **Final Efficacy Analysis Randomized Population** consists of all patients who were randomized prior to or on August 4, 2020.
- **Safety Population** consists of all randomized patients who received at least 1 dose of study drug. Safety analysis will be based on the Safety Population, in which patients are evaluated according to the treatment they received.
- **Modified Intent-to-Treat (mITT) Population** consists of all patients who were randomized prior to or on August 4, 2020, and who received at least 1 dose of study drug and had at least 1 postbaseline primary efficacy assessment. This population will be used for the final efficacy analysis.
- **Sensitivity Analysis (SA) Population** consists of all randomized patients who received at least 1 dose of study drug and had at least 1 postbaseline primary efficacy assessment. This population will be used for the efficacy sensitivity analysis at completion of the study.

7.2. Patient Disposition

The number and percentage of screen failure patients (ie, patients who entered the Screening Period but not the Pretreatment Period), the number and percentage of pretreatment failures (ie, patients who entered the Pretreatment Period but were not randomized), along with the associated reasons for failure, will be tabulated overall for the Screened Population. Patients who initially failed screening and were re-screened will only be counted once, that is, the PID used for the rescreened patients will be listed for patient disposition, while the PID assigned for the original screening will not be analyzed or listed.

The number and percentage of randomized patients who were included in the Safety Population, who were included in the mITT Population, who completed the Treatment Period, and who prematurely discontinued treatment will be presented overall and by treatment group. A patient is considered to have completed the Treatment Period if the patient completed the 8 weeks of treatment and the EOT Visit that was recorded on the end-of-treatment eCRF page. The reason for premature discontinuation as recorded on the eCRF will also be summarized by treatment group and overall.

The number and percentage of patients who prematurely discontinued treatment will be presented for each treatment group by time of discontinuation, defined as the week (Week 1 through Week 8) in which the treatment discontinuation occurred.

The number of patients screened, the number and percentage of patients randomized will be presented by study center within each geographic region (See Appendix 3 for the definitions of

the geographic regions). A separate disposition table will be generated to include the SA Population and other relevant information.

7.3. Demographics and Other Baseline Characteristics

Demographic parameters (age, age group, race, ethnicity, sex), baseline characteristics (weight, height, and BMI that calculated as weight in kg/[height in m]^2), will be summarized descriptively for the Safety, mITT and SA Populations by treatment group and overall.

Baseline efficacy parameters and randomization stratification variables (including EE status, WHSS and its category [3 vs. ≥3], WRFS, proportion of heartburn-free days, the weekly score of each mRESQ item, sleep disturbance, symptom bothersomeness, and degree of symptom relief) will be summarized descriptively for the mITT and SA Populations by treatment group.

There are several different data derivations for baseline WHSS (Section 5.2.1) and EE status (Section 5.4), the baseline values for each of them will be reported separately for the mITT and SA Populations.

7.4. Compliance

7.4.1. Treatment Compliance

Compliance with study drug dosing (IW-3718 or matching placebo) is defined as the total number of tablets taken by a patient during the study divided by the number of tablets that were expected to be taken, multiplied by 100. The total number of tablets taken will be calculated based on the following formula:

total number of tablets dispensed – (number of tablets returned + number of tablets lost)

The number of tablets lost is recorded on eCRF. The total number of tablets expected to be taken during the study equals the number of days in the study times 6 (minus 3 if the date ends on the day of EOT Visit).

The study drug compliance data will be reported using descriptive statistics for the Safety Population by treatment group. The compliance data will be categorized for increment of 20%, and it will be presented using patient counts and percentages for the Safety Population by treatment group. Additionally, compliance to PPI for the Pretreatment Period and for the Treatment Period will be summarized respectively similarly as the study drug.

7.4.2. eDiary Compliance

Patients are required to enter their diary assessments into the eDiary before bed each night, except assessment of sleep disturbance that is assessed upon getting up each morning during the Pretreatment and Treatment Periods. A daily report is considered complete if all questions in both the morning diary and evening diary (excluding the weekly assessments) are answered, with the following exceptions:

- For the first day of the Pretreatment Period, only the evening diary is expected.
- For the last day of the Pretreatment Period, only the morning diary will be used.
- For the first day of the Treatment Period, only the evening diary will be used.
- For the EOT Visit day, only the morning diary is expected.

eDiary compliance will be calculated for baseline (the Pretreatment Period), the Treatment Period, and weekly within the Treatment Period. For each period, eDiary compliance will be calculated as the number of completed daily reports divided by the expected number of daily reports in the period multiplied by 100.

The eDiary compliance data will be reported using descriptive statistics for the mITT Population by treatment group. The compliance data will be classified by increment of 20%, and it will be presented using patient counts and percentages for the mITT Population by treatment group. Additionally, categorized (≥4 vs. <4) weekly eDiary data will be presented using patient counts and percentages by treatment group for the mITT Population.

7.5. Extent of Exposure

Exposure to study drug will be calculated as the number of days from the date of first dose taken to the date of last dose taken in the Treatment Period:

Treatment Duration = Date of last dose of study drug - Date of first dose of study drug + 1

Exposure to study drug for the Safety Population during the Treatment Period will be summarized descriptively by treatment group. Treatment duration will also be categorized as (in days):

- 1
- >1 to ≤7
- >7 to ≤14
- >14 to ≤21
- >21 to ≤28
- >28 to ≤35
- >35 to ≤42
- >42 to ≤49
- >49 to ≤56 and
- >56

The number and percent of patients in each category will be presented by treatment group for the Safety Population.

Patient-years, defined as total exposure days to the study drug across all patients and divided by 365.25, will be summarized for the WI-3718 group for the Safety Population.

Additionally, exposure to PPI for the Pretreatment Period and for the Treatment Period will be summarized, respectively, similarly to the study drug.

7.6. Protocol Deviations

Protocol deviations and Important Protocol Deviations (IPD) will be identified and documented for all randomized patients prior to unblinding through programmatic checks of the study data and select individual data reviews. IPDs will be determined based on a blinded review of all protocol deviations and protocol deviation categories prior to database lock and unblinding; this review will be performed by members of the Ironwood Clinical Trial team, including but not limited to the Study Program Manager, Clinical Medical Monitor. and Study Biostatistician.

The number and percentage of patients with IPDs will be presented by treatment group and IPD category for the Randomized Population. Protocol deviations and IPDs will be provided by patient in data listings for the Randomized Population.

A separate summary and listing of all inclusion/exclusion criteria deviations will also be provided. This summary will be based on data as recorded on the inclusion/exclusion page of the eCRF.

7.7. Prior and Concomitant Medication

Patient's daily eDiary recording of PPI administration will not be reported as a prior or concomitant medication. Instead, it will be summarized separately. Details are specified in Section 7.4.1 and Section 7.5.

Prior medication is defined as any medication started before the date of the first dose of study drug.

Concomitant medication is defined as any medication with a start date prior to, or on the date of the first dose of study drug and continuing after the first dose of study drug, or with a start date between the dates of the first and last doses of study drug, inclusive. Any medication with a start date after the date of the last study drug will not be considered a concomitant medication.

Both prior and concomitant medications will be coded by drug name and therapeutic class using WHO Drug dictionary B2 format, March 2018 version or above. The use of prior and concomitant medications will be summarized for the Safety Population as the number and percentage of patients in each treatment group receiving each medication within each therapeutic class. If a patient took a specific medication multiple times or took multiple medications in the same category (based on Anatomical-Therapeutic-Chemical classification), that patient would be counted only once for the coded drug name or therapeutic class. The use of per-protocol rescue medicines is collected separately in the eDiary (Section 7.8).

7.8. Rescue Medicine Use

During the Pretreatment and Treatment Periods, patients may use dispensed, protocol-permitted rescue medicine when their heartburn becomes intolerable. Patients will report the frequency of their use of rescue medicine (liquid antacid) in the evening eDiary before bed for the day. Overall daily use of rescue medicine will be summarized for each treatment week by treatment group for the Safety Population.

7.9. Efficacy Analysis

The final efficacy analysis will be based on the mITT, while the SA Population will be used for the sensitivity efficacy analysis, and the analyses will be conducted according to the treatment assigned. Baseline values for efficacy parameters are defined as:

- For change from baseline continuous parameter and derived dichotomized response variables, the baseline will be derived from the eDiary and/or eCRF data collected for the last week of the Pretreatment Period (or Week -1), specifically the period from 7 days prior to the day of randomization.
- A time dependent proportion efficacy parameter that is defined as days the event of interest occurs, divided by the number of diary entry days during the relevant period. Its baseline period is the last 14 days (Week -2 to Week -1) prior to day of randomization.

For the change from baseline and response variables, if data for Week -1 are not available, data for Week -2 (the second last week of the Pretreatment Period) will be used.

For responder parameters derived from daily eDiary assessments, a patient who missed 4 or more daily assessments an analysis week will be consider a missing for that week and will be defined as a non-responder for that week. For continuous efficacy parameters, the analysis value for an analysis period is defined as the average of the non-missing values during that period (Appendix 1). If a patient who reported assessments for fewer than 1 day during a week within the Treatment Period, the average is missing for that week. For efficacy parameters no further imputation will be performed (see Section 7.9.1.3 for sensitivity analyses employing multiple imputation [MI] methods). Any efficacy assessments measured after the last dose date will not be included for analysis.

All primary and secondary efficacy parameters are listed in Table 4 will be tested in the given order to protect the overall family wide alpha level at 5%. The mITT Population will be used for the final analysis for all efficacy parameters. The SA Population will be used for the efficacy sensitivity analysis that include selected efficacy parameters at completion of the study.

7.9.1. Primary Efficacy Parameter

7.9.1.1. Estimands

The primary efficacy estimands that are prepared to support regulatory decisions, are described below.

The **While on Treatment Strategy** for the primary efficacy parameter will be used. It is defined by the following:

- **Treatment**: intervention IW-3718 (1500 mg), while receiving PPIs.
- **Population**: patients in the mITT with GERD who continue to have persistent symptoms, such as heartburn and regurgitation, while receiving once-daily (QD), standard-dose PPIs.

- **Variable**: Weekly Heartburn Severity Score (WHSS) derived for each treatment visit. The WHSS for a week is the average of the non-missing Daily Heartburn Severity Scores (DHSS) for that week.
- **Intercurrent event**: the following intercurrent events will be considered:
  a. Discontinuation of treatment due to lack of efficacy
  b. Lost to follow-up
  c. Discontinuation of treatment due to AE
  d. Discontinuation in the study by physician's decision
  e. Withdrawing consent
  f. Withdraw due to protocol violations
- **Population-level summary**: mean difference in change from baseline between treatment groups.

Henceforth, for simplicity, background therapy PPIs will be omitted when referring to study drug treatment (IW-3718 and placebo).

7.9.1.2. Analysis of the Primary Efficacy Parameter

The primary efficacy parameter of the study is defined as the change from baseline in WHSS at Visit 5 (Week 8):

- The WHSS for a week is the average of the non-missing Daily Heartburn Severity Scores (DHSS) for that week (Appendix 1).
- The DHSS for a day is the greater score of the 2 mRESQ-eD items assessing heartburn severity ("Burning feeling behind the breastbone or in the center of the upper stomach" and "Pain behind the breastbone or in the center of the upper stomach") for that day. Where one of the 2 mRESQ-eD heartburn severity items is missing for the day, the maximum will be the score of the remaining item; where if both of the items are missing for the day, the DHSS will be missing.

The primary efficacy parameter will be analyzed using a linear mixed-effects model for repeated measures (MMRM). The model will include WHSS change from Baseline at postbaseline scheduled visits up to Week 8 as dependent variable; treatment, randomization stratification variables Baseline EE status (present versus not present) and Baseline WHSS severity level (<3 versus ≥ 3), Week (as a categorical variable), Week-by-treatment and Week-by-Baseline WHSS interactions as fixed effects; Baseline WHSS as a continuous covariate; and subject as a random effect. The Kenward-Roger approximation will be used to estimate denominator degree of freedom and adjust standard errors. An unstructured covariance matrix will be used for the MMRM analysis. In the case that the convergence is an issue, the compound symmetry covariance structure will be used. Least square (LS) mean for each treatment group, LS mean difference between the IW-3718 group and the placebo group and a 95% CI for the difference, as well as the *P* value for comparison versus placebo will be presented. For this analysis, the primary contrast of interest is at Visit 5 (Week 8) for IW-3718 compared with placebo.

A corresponding line graph of LS mean with 95% CI by treatment group and Week will be presented. Cumulative distribution function (CDF) of change from baseline at Week 8 in WHSS to be plotted by treatment group. To aid in the interpretation of the graphical representation of the CDF across treatment group, a 2-sample Kolmogorov-Smirnov test will be conducted. If the *P*-value of the Kolmogorov-Smirnov test is small, conclude that the two treatment groups were sampled from populations with different distributions. The populations may differ in median, variability, or the shape of the distribution.

CDF plots of change from Baseline at each week in WHSS by treatment group and patient-reported weekly Degree of Relief Assessments (comparing the symptoms over the past 7 days, Appendix 7) classified anchor will be provided. For a sense of variability in the distribution of the classified heartburn relief anchor category, percentile (10^{th}, 25^{th}, 75^{th} and 90^{th}) of change from Baseline to each week in WHSS by treatment group and the anchor will be tabulated.

7.9.1.3. Sensitivity Analysis Based on Multiple Imputation – Missing Not at Random (MNAR)

The analysis specified in this section will be executed only when a superiority result that favors the active treatment group for the primary efficacy parameter is obtained.

To characterize the extent of missing data associated with the primary efficacy parameter, the number and percentage of patients with missing evening diaries will be presented by treatment group for the mITT Population for each week of the Treatment Period.

The primary efficacy analysis (MMRM) relies on the assumption that the missing data mechanism follows the missing at random (MAR) scenario. In this case, it is assumed that the reason for data being missing may depend on observed data, but not on the unobserved missing data. The likelihood based MMRM analysis is an appropriate method for the statistical analysis under MAR assumption.

The following sensitivity analysis model, which are within the pattern-mixture model framework, will be used to examine the robustness of the primary efficacy analysis results for the missing not at random (MNAR) mechanisms. Under MNAR, it is assumed that the reason for data being missing is related to the unobserved missing data.

The imputations are based on the distribution of placebo group responses over time. The underlying assumption is that the missing data for a patient on the active treatment follow the distribution of the placebo responses, i.e., the mean values and intra-subject correlations based on the placebo responses will be applied. Moreover, the intercurrent events will generate MNAR (informative) missing, which will have impact on determining of drug effect. A designated penalty will be prespecified for this type of missing mechanism, in order to adjust the treatment effect estimation to determine a possible “true” treatment effect.

The model is implemented in 4 steps:

Step 1: Imputations

For change from baseline data, a total of 200 sets of posterior mean and co-variance estimates are extracted from the available non-missing placebo data. One hundred of the posterior sets will be applied to the pooled active treatment group, while the other 100 will be applied to the placebo group. One set of imputations for all missing values will be generated based on each variation of posterior estimates. All 100 datasets for imputations within a treatment group will be

ordered from 1 to 100 and combined between pooled active treatment group and placebo, for a total of 100 completely imputed datasets.

Step 2: Adjustments

For patients who have any intercurrent events, a penalty will be designated to the imputed values at the last scheduled visit (Visit 5, Week 8). The penalty is applied to the 2 treatment arms by adding a fraction of the estimated SD for the primary efficacy parameter (the square root of the estimated element for the last scheduled visit of the co-variance matrix R from the primary MMRM model): (0×SD), (0.25×SD), (0.50×SD).

Step 3: Analysis of Complete Datasets

The primary efficacy parameter will be analyzed for each of the 100 complete datasets with imputed and the designated penalty applied data at the last scheduled visit (Visit 5, Week 8), using an analysis of covariance with treatment group as factors and the baseline value as a covariate.

Step 4: Inference

Rubin's Rules will be applied to pool the desired statistical quantities that include LS mean difference estimates, associated 95% CI and p-value, and associated SEs.

Efficacy conclusion of the study is based on the analysis described in Section 7.9.1.2. In the current section, a different missing mechanism is assumed for the data, it is a supportive outcome for the primary efficacy effect analysis.

7.9.2. Key Secondary Efficacy Parameters

There are 3 key secondary efficacy parameters which will be included in the hypothesis testing procedure.

1. Change from Baseline at Week 8 in WRFS
    - The WRFS is an average of the non-missing Daily Regurgitation Frequency Scores (DRFS) for that week (Appendix 1).
    - The DRFS for a day is the greater score of the 2 mRESQ-eD items assessing regurgitation frequency ("Regurgitation [liquid or food moving upwards toward your throat or mouth]" and "An acid or bitter taste in the mouth") for that day. Where 1 of the 2 mRESQ-eD regurgitation frequency items is missing for the day, the maximum will be the score of the remaining item; where both of the items are missing for the day, the DRFS will be missing.
2. Overall Heartburn Responder (OHR)
    - An OHR is a patient who is a Weekly Heartburn Responder for at least 4 weeks, including at least 1 of the last 2 weeks (Weeks 7 and 8), during the 8-week Treatment Period.
    - A patient who reports a DHSS for less than 4 days during a week will be considered a non-responder for that week.
    - A Weekly Heartburn Responder is a patient with a decrease of ≥45% from baseline in WHSS.

3. Proportion of heartburn-free days during the 8-week Treatment Period

- Proportion of heartburn-free days is calculated as the number of heartburn-free (DHSS = 0) days divided by the number of diary entry days (Appendix 1).

7.9.2.1. Analysis of the Secondary Efficacy Parameters

If the primary efficacy comparison generates a superiority result for IW-3718, the secondary efficacy parameters will be tested and so on. The testing order and analysis methods are listed below:

Table 6: Analysis and Assumptions of the Secondary Efficacy Parameters

Order of Hypothesis Testing	Parameters	Analysis Method and Outputs	Assumptions
1	Change from Baseline at Week 8 in WRFS	Efficacy parameter WRFS will be analyzed using the MMRM approach like the primary efficacy parameter, WHSS change from baseline.	MAR
2	Overall Heartburn Responder	The proportions of responders between the IW-3718 group and the placebo group will be compared using a Cochran-Mantel-Haenszel (CMH) test controlling for baseline EE status (present vs. not present), baseline WHSS (<3 vs. ≥3). The CMH test is the primary analysis for responder parameters. The estimated odds ratio (IW-3718 over placebo) associated 95% CI and *P* value based on CMH test will be tabulated. Additionally, difference in responder rate with a 95% Newcombe CI (3) will also be provided.	Missing will not be imputed.
3	Proportion of Heartburn-free Days During the 8-week Treatment Period	The data will be analyzed using the Poisson regression, which is appropriate for count data. The analysis includes the fixed categorical effect of treatment, the baseline esophagitis status (present vs. not present) and baseline WHSS (<3 vs. ≥3), the covariate of baseline proportion of heartburn-free days, and with the diary entry duration (in days) as a weight variable[a] adjusted in the model. Model estimates in difference of incidence rate between the IW-3718 treatment group and placebo will be represented with corresponding 95% CI and *P*-value associated with the comparisons to placebo. In the case of overdispersion, a negative binomial model instead of Poisson regression will be implemented.	Data following a Poisson distribution. Missing data will not be imputed.

a. The weight variable weights the overdispersion parameter, which has the default value of 1. The weight variable does not have to be an integer, if it is less than or equal to 0 or if it is missing, the corresponding observation is not used for modeling analysis.

Like the primary efficacy parameter, a corresponding line graph of LS mean with 95% CI by treatment group and Week for WRFS will be presented. CDF of change from baseline at Week 8 in WRFS to be plotted by treatment group. To aid in the interpretation of the graphical representation of the CDF across treatment group, a 2-sample Kolmogorov-Smirnov test will be conducted.

CDF plots of change from Baseline at each week in WRFS by treatment group and patient-reported weekly Degree of Relief Assessments (comparing the symptoms over the past 7 days, Appendix 7) classified anchor will be provided. For a sense of variability in the distribution of the classified heartburn relief anchor category, percentile (10^{th}, 25^{th}, 75^{th} and 90^{th}) of change from Baseline to each week in WRFS by treatment group and the anchor will be tabulated.

7.9.3. Exploratory Efficacy Parameters

Exploratory efficacy parameters will be explored outside of the formal testing procedures and not controlled for multiplicity. The mITT Population will be used for the analyses. If appropriate, for selected efficacy parameters, the SA Population will be used for sensitivity analysis. The baseline definitions for different type of parameters can be found in Section 7.9.

- Change or percent change continuous parameters will be analyzed using the MMRM approach like the primary efficacy parameter, WHSS change from baseline.
- Categorical responder parameters will be analyzed utilizing the CMH testing method that is proposed for the secondary efficacy analysis for OHR.
- Time dependent proportion parameters, defined as days the event of interesting occurs divided by the number of eDiary entry days during the relevant period, will be analyzed applying Poisson molding, which is appropriate for count data. The analysis is similarly as the analysis for the proportion of the heartburn-free days during the 8-week Treatment Period.

The exploratory efficacy parameters and proposed analysis methods are listed below:

Table 7: Exploratory Efficacy Parameters and Analysis Methods

Item No.	Efficacy Endpoints	Analysis Methods
1.	Percent change from baseline at Week 8 in WHSS	MMRM for change from baseline to Week 8.
2.	Overall Heartburn Responders with baseline WHSS ≥3	CMH
3.	Change from baseline of proportion of days when heartburn did not occur (DHSS = 0) or was very mild (DHSS = 1) during the 8-week Treatment Period	Analysis of covariance (ANCOVA)
4.	Proportion of days when heartburn did not occur (DHSS=0) or was very mild (DHSS=1) during the 8-week Treatment Period	Poisson regression for proportions
5.	Percent change from baseline at Week 8 in WRFS	MMRM for change from baseline to Week 8.

Item No.	Efficacy Endpoints	Analysis Methods
6.	Change from baseline proportion of days when regurgitation did not occur (DRFS = 0) or rarely occurred (DRFS = 1) during the 8-week Treatment Period	ANCOVA
7.	Proportion of regurgitation free (DRFS = 0) days during the 8-week Treatment Period	Poisson regression for proportions
8.	Proportion of days when regurgitation did not occur (DRFS=0) or rarely occurred (DRFS=1) during the 8-week Treatment Period	Poisson regression for proportions
9.	Overall Regurgitation Responder: a patient who is a Weekly Regurgitation Responder for at least 4 weeks, including at least 1 of the last 2 weeks (Weeks 7 and 8), during the 8-week Treatment Period • A Weekly Regurgitation Responder is a patient with a decrease of ≥45% from baseline in WRFS • A patient who reports DRFS for less than 4 days during a week will not be considered a Weekly Regurgitation Responder for that week	CMH
10.	Change from baseline to Week 8 in each mRESQ item	MMRM for change from baseline to Week 8.
11.	Proportion of nights with difficulty falling asleep due to GERD during the 8-week Treatment Period	Poisson regression for proportions
12.	Proportion of days with nighttime awakening (≥1 time) due to GERD during the 8-week Treatment Period	Poisson regression for proportions
13.	Degree of relief of heartburn/regurgitation/overall GERD symptoms responder: • Degree of relief responder: patient who reported a score of 1-Significantly Relieved or 2-Moderately Relieved on the Degree of Relief question for at least 4 weeks, including at least 1 of the last 2 weeks (Weeks 7 and 8), during the 8-week Treatment Period	CMH
14.	Change from baseline at Week 8 in heartburn/regurgitation/overall GERD symptoms bothersomeness	MMRM for change from baseline to Week 8.
15.	Proportion of days with rescue medication use during the 8-week Treatment Period	Poisson regression for proportions
16.	Erosive esophagitis (EE) healed (not present) at Week 8	CMH (the central reads and the local reads will be analyzed, respectively)

Item No.	Efficacy Endpoints	Analysis Methods
17.	Proportion of patients with erosive esophagitis (EE) improved by at least 1 grade at Week 8	CMH (the central reads and the local reads will be analyzed, respectively)

Descriptive supportive summary for following variables will be generated for SA population.

Item No.	Description of Summary
1.	Percent change from baseline in WHSS by week
2.	Percent change from baseline in WRFS by week
3.	Change from baseline in WHSS by week
4.	Change from baseline in WRFS by week
5.	Change from baseline in each mRESQ item by week
6.	Proportion of heartburn-free (DHSS=0) days by week
7.	Proportion of regurgitation-free (DRFS=0) days by week
8.	Proportion of days when heartburn did not occur (DHSS=0) or was very mild (DHSS=1) by week
9.	Proportion of days when regurgitation did not occur (DRFS=0) or rarely occurred (DRFS=1) by week
10.	Proportion of nights with difficulty falling asleep due to GERD by week
11.	Proportion of days with nighttime awakening (≥1 time) due to GERD by week
12.	Proportion of days with rescue medication use by week
13.	Degree of relief of heartburn/regurgitation/overall GERD symptoms by week
14.	Heartburn/regurgitation/overall GERD symptoms bothersomeness by week
15.	Treatment satisfaction by week
16.	Shift in erosive esophagitis (EE) grade from baseline to Week 8

7.10. Safety Analysis

All safety parameters will be summarized using descriptive statistics. Safety analyses will be performed on the Safety Population. The safety parameters will include AEs, clinical laboratory evaluations, vital signs, ECGs, and physical examination. For each safety parameter, the last non-missing assessment made before the first dose of study drug will be used as the baseline for all analyses of that safety parameter.

The treatment group designation for safety summaries is based on the treatment the patient actually received. The “treatment actually received” for a patient who received treatment other

than the assigned treatment (even if only a single dose) during the treatment period will be defined as the treatment of which the patient actually took, regardless of the treatment assigned.

7.10.1. Adverse Events

AEs will be coded by system organ class and preferred term using the MedDRA dictionary, version 23.0 which contains the COVID-19 specific terms. An AE will be considered a treatment-emergent AE (TEAE) if the AE started after initial study drug administration and within 1 day of the last dose of study drug. In addition, an AE that started before initial study drug but worsened after the first dosing of the study drug is also considered a TEAE.

The number and percentage of patients reporting TEAEs in each treatment group will be tabulated:

- by system organ class (SOC) and preferred term (PT)
- by SOC, PT, and severity
- by SOC, PT, and relationship to study drug.

If a patient has more than one TEAE coded to the same PT, the patient will be counted only once for that PT by identifying those TEAEs with the highest severity and the closest relationship to study drug.

The incidence of the following TEAEs will be summarized by PT:

- Most frequent (≥2% of patients in any treatment group) TEAEs
- Treatment Emergent SAEs
- TEAEs leading to premature discontinuation of study drug.

All AE summary tables will be SOC sorted alphabetically and have PT within each SOC sorted by decreasing frequency for IW-3718 group.

A TEAE summary table with all the non-SAEs by SOC and PT will be generated to support the clinicaltrials.gov reporting purpose.

AE listings will be presented for all patients with AEs, screened patients with SAEs, patients with AEs leading to discontinuation, all SAEs (no matter it is treatment-emerge or not) and patients who died (if any), respectively.

7.10.2. Clinical Laboratory Parameters

Descriptive statistics for clinical laboratory values (in SI units) and changes from the baseline values at each assessment time point will be presented by treatment group for each clinical laboratory parameter. All laboratory data will be listed. The clinical laboratory parameters are listed in Section 5.5.6.

For each clinical laboratory parameter, the number and percentage of patients with potentially clinically significant (PCS) postbaseline clinical laboratory values will be tabulated by treatment group for data collecting no more than 1 day after the last dose. The criteria for PCS laboratory values are described in Table 8. The percentages will be calculated relative to the number of patients with available non-PCS baseline values and at least 1 post-baseline assessment. The

potential numerator will be the total number of patients with available non-PCS baseline values and at least 1 post-baseline PCS value in the corresponding postbaseline period.

A supportive listing for patients with PCS postbaseline values will be provided, including the PID number, study center, baseline and postbaseline values for each visit, and a flag indicating the PCS laboratory values.

Table 8: Criteria for Potentially Clinically Significant Laboratory Results

Parameter	SI Unit	Lower Limit	Higher Limit
CHEMISTRY			
Albumin	g/L	<0.9 × LLN	>1.1 × ULN
Alanine aminotransferase	U/L	—	≥3 × ULN
Alkaline phosphatase	U/L	—	≥3 × ULN
Aspartate aminotransferase	U/L	—	≥3 × ULN
Bicarbonate	mmol/L	<0.9 × LLN	>1.1 × ULN
Bilirubin, total	µmol/L	—	>1.5 × ULN
Calcium	mmol/L	<0.9 × LLN	>1.1 × ULN
Chloride	mmol/L	<0.9 × LLN	>1.1 × ULN
Cholesterol, total	mmol/L	—	>1.6 × ULN
Creatinine	µmol/L	—	>1.3 × ULN
Glucose	mmol/L	<0.8 × LLN	>1.4 × ULN
Magnesium	mmol/L	<0.9 × LLN	>1.1 × ULN
Phosphate	mmol/L	<0.9 × LLN	>1.1 × ULN
Potassium	mmol/L	<0.9 × LLN	>1.1 × ULN
Protein, total	g/L	<0.9 × LLN	>1.1 × ULN
Sodium	mmol/L	<0.9 × LLN	>1.1 × ULN
Urea nitrogen	mmol/L	—	>1.2 × ULN
Uric acid	µmol/L	<0.9 × LLN	>1.1 × ULN
HEMATOLOGY			
Basophils, absolute cell count	10^9/L	—	>3 × ULN
Eosinophils, absolute cell count	10^9/L	—	>3 × ULN
Hematocrit	Ratio	<0.9 × LLN	>1.1 × ULN

Parameter	SI Unit	Lower Limit	Higher Limit
Hemoglobin	g/L	<0.9 × LLN	>1.1 × ULN
Lymphocytes, absolute cell count	10^9/L	<0.8 × LLN	>1.5 × ULN
Mean corpuscular hemoglobin	Pg	—	>3 × ULN
Mean corpuscular hemoglobin concentration	g/L	—	>3 × ULN
Mean corpuscular volume	fL	<0.9 × LLN	>1.1 × ULN
Monocytes, absolute cell count	10^9/L	—	>3 × ULN
Neutrophils, absolute cell count	10^9/L	<0.8 × LLN	>1.5 × ULN
Platelet count	10^9/L	<0.5 × LLN	>1.5 × ULN
Red blood cell count	10^{12}/L	<0.9 × LLN	>1.1 × ULN
White blood cell count	10^9/L	<0.7 × LLN	>1.5 × ULN

LLN = lower limit of normal value provided by the laboratory; SI = *Le Système International d'Unités* (International System of Units); ULN = upper limit of normal value provided by the laboratory.

7.10.3. Vital Signs

Descriptive statistics for body weight and vital signs (ie, oral temperature, respiratory rate, systolic and diastolic blood pressure, and pulse rate) and changes from baseline values at each visit will be presented by treatment group, and all the data will be listed.

The number and percentage of patients with PCS postbaseline vital signs will be tabulated by treatment group for data collecting no more than 1 day after the last dose. A vital sign value will be considered PCS if it meets both the observed value criterion and the change from baseline value criterion. The criteria for PCS vital sign values is detailed in Table 9. The percentages will be calculated relative to the number of patients with baseline values and at least 1 assessment in the corresponding postbaseline period. The numerator will be the total number of patients with available baseline values and at least 1 PCS value in the corresponding postbaseline period.

A supportive listing for patients with PCS postbaseline values will be provided, including the PID number, study center, baseline and postbaseline values for each visit, and a flag indicating the PCS vital sign values.

Table 9: Criteria for Potentially Clinically Significant Vital Signs

Parameter	Flag	Criteria[a]	
		Observed Value	Change from Baseline
Seated systolic blood pressure, mmHg	High	≥180	Increase of ≥20
	Low	≤90	Decrease of ≥20
Seated diastolic blood pressure, mmHg	High	≥105	Increase of ≥15
	Low	≤50	Decrease of ≥15
Seated pulse rate, bpm	High	≥120	Increase of ≥15
	Low	≤50	Decrease of ≥15
Weight, kg	High	—	Increase of ≥7%
	Low	—	Decrease of ≥7%

a. A postbaseline value is considered potentially clinically significant if it meets both the criteria for observed value and the criteria for change from baseline.

bpm = beats per minute

7.10.4. ECG Parameters

Descriptive statistics for ECG variables (eg, QRS duration, QT interval, PR interval, and QTcF interval.) and their changes from baseline at each assessment time point will be presented by treatment group. ECG interpretation will be summarized by visit. A shift table from baseline to each visit for qualitative ECG results will be presented.

In addition, the number and percentage of patients with PCS postbaseline ECG will be tabulated by treatment group for data collecting no more than 1 day after the last dose. An ECG value will be considered PCS if it meets the observed value criterion. The criteria for PCS ECG values is detailed in Table 10. The percentages will be calculated relative to the number of patients with available non-PCS baseline values and at least 1 postbaseline assessment. The potential numerator will be the total number of patients with available non-PCS baseline values and at least 1 postbaseline PCS value in the corresponding post-baseline period.

A supportive listing for patients with PCS postbaseline values will be provided, including the PID number, study center, baseline and postbaseline values for each visit, and a flag indicating the PCS ECG values.

Table 10: Criteria for Potentially Clinically Significant ECG Parameters

ECG Parameter	Unit	Higher Limit
QRS duration	msec	≥150
PR interval	msec	≥250
QTcF interval	msec	>500

7.10.5. Medical and Disease History

Abnormalities in patients' medical and surgical histories will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), version 21.0. The number and percentage of patients with abnormalities in medical and surgical histories in each system organ class (SOC) and preferred term will be presented by treatment group and overall, for the Safety Population.

7.10.6. Physical Examination

No separate analysis for physical examinations is planned; a listing will be provided for physical examination findings.

7.11. Health Outcome Analysis

Health outcome analyses will be performed on the SA Population.

7.11.1. WPAI-Sleep Disturbance-GERD

Change from baseline in hours missed from work due to sleep disturbance from GERD as a percentage of total hours (the summation of hours missed due to sleep disturbance from GERD, hours missed due to any other reason, and hours worked) will be summarized by treatment group at Weeks 4 and 8 and analyzed using a similar MMRM approach described in Section 7.9.1.2.

Change from baseline in the effect of sleep disturbance due to GERD on work productivity and sleep disturbance due to GERD on regular daily activities other than work at a job will be summarized, respectively, by treatment group at Weeks 4 and 8 and analyzed using a similar MMRM approach described in Section 7.9.1.2.

7.11.2. EQ-5D-3L

Each of the 5 questions assessing mobility, self-care, usual activities, pain/discomfort, and anxiety/depression will be summarized by treatment group and by frequency distribution (profile) of the EQ-5D descriptive system ("No problems", "Some problems", "Extreme problems") at baseline, Week 4 and Week 8. Comparison between treatment groups at Week 8 will be performed utilizing the CMH test, controlling for strata: baseline EE status (present vs. not present), baseline heartburn severity level (<3 vs. ≥ 3) and the associate p-value will be reported. In addition, shift in response to each question from baseline to each visit postbaseline will be summarized using a shift table.

Change from baseline at week 4 and Week 8 in health state as assessed by a 0 to 100 visual analogue scale will be summarized by treatment group and analyzed using a similar MMRM approach described in Section 7.9.1.2.

8. REFERENCES

1. Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency Guidance for Industry, U.S. Department of Health and Human Services Food and Drug Administration (June, 2020)

2. EuroQol G. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy. 1990; 16:199-208.

3. Newcombe, R. Interval estimation for the difference between independent proportions: Comparison of eleven methods. Stat. Med. l 998; 17: 873-890.

4. Cohen, J. *Statistical power analysis for the behavioral sciences,* Hillsdale, NJ, Lawrence Earlbaum Associates. 1988

5. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care. 2005;43(3):203-20.

6. Ratitch, B., O'Kelly, M. and Tosiello, R. (2013), Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. Pharmaceut. Statist., 12: 337–347. doi: 10.1002/pst.1549

7. Rubin DB (1987) Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons Ltd., New York.

Appendix 1. Missing Data Simulation to Support the Key Efficacy Parameters Scoring Algorithm

A missing data simulation study was conducted to assess the impact of the missing data on the scoring algorithm of the key efficacy parameters. The following assumptions were made for the simulation:

- Patients with evidence of pathological acid reflux (positive Bravo status) at baseline in Study ICP-3718-202 represent a study population that is similar to the current study population.
- The efficacy parameters WHSS, WRFS and PHFD in the current study follow the same underlying data distribution as in Study ICP-3718-202.

In Study ICP-3718-202, 45 patients with a positive Bravo status at baseline had non-missing daily assessments in the placebo and IW-3718 1500 mg BID groups combined. The daily assessments were used to derive the efficacy parameters WHSS, WRFS and PHFD. These 45 patients were used for the simulation study.

In Study ICP-3718-202, approximately 10% of patients prematurely discontinued from the study. In the initial simulation, it was assumed that 10% of patients had a given number of missing daily data at each week (ie, the missing data rate). In a second simulation, it was assumed that 20% of patients had a given number of missing daily data at each week.

Daily Scores

Derivation of PHFD is based on Item #1 of mRESQ-eD (Appendix 5). The simulations were run on the value of Item #1.

The daily score that is used to derive the WHSS is the greater score of Item #1 and Item #2 of the mRESQ-eD, which both assess heartburn severity for a given day.

Similar to the WHSS, the daily score that is used to derive the WRFS is the greater score of Item #6 and Item #7 of the mRESQ-ed, which both assess regurgitation frequency for a given day.

It was assumed that a scenario whereby a patient misses 1 of the 2 mRESQ-eD items does not occur. The reason is that when a patient answers a mRESQ-eD question, the next question automatically appears on the device, and it cues the patient to answer that question. Given the nature of the device, there are most likely only 2 scenarios for the mRESQ-eD data collection: 1) a patient answers all mRESQ-eD items; 2) none of the items are answered. The experience of Study ICP-3718-202 supports this. There was no instance where the eDiary had partial data for a given day. Similarly, this has been the experience in the 2 ongoing Phase 3 studies.

Based on the above discussions, the simulations conducted for WHSS and WRFS began with the daily scores, rather than the item scores. The WHSS and WRFS are calculated as the weekly average of the non-missing daily scores.

Weekly Heartburn Severity Score and Weekly Regurgitation Frequency Score

For the weekly scoring algorithm, a minimum number of non-missing daily scores over the 7-day period should be required to calculate the weekly scores. In the simulation study, we randomly replaced non-missing daily scores with missing daily scores for an increasing number

of days, beginning with 1 day and proceeding up to 6 days. We also ran the simulations with 2 different percentages of patients with missing data at each week, 10% and 20%.

A patient cluster is defined as all patient daily scores with a weekly repeated measures structure. Given the sample size of the current study, approximately 550 patients, and a patient cluster bootstrapping resampling based on Study ICP-3718-202 subset that includes 45 patients, we performed the following steps for the simulation:

1. A bootstrapping resampling process with replacement to select 550 patient clusters with 2000 replications was applied.
2. In each replicated dataset, a given percentage (eg, 10% or 20%) of patient clusters was randomly selected.
3. In each replicated dataset, in each patient cluster that was selected for in Step 2, at each week, a given number of daily scores (ie, 4, 5, or 6 days) was randomly selected and set to missing. (Each patient has a planned 8 weeks of data assessments and at each week there are 7 planned daily scores.)
4. Average change from baseline by week was calculated for both non-missing and missing data, respectively.

Across all replicated datasets, for both the data without any missing daily scores and the data replaced by daily missing scores, their resultant weekly average score, median, and standard deviation were calculated. Comparing the summary scores between these 2 data conditions helped us determine the minimum number of non-missing daily scores that should be included in the efficacy parameter scoring algorithms.

The summary statistics for the efficacy parameter WHSS are presented in Table 11. The descriptive summary statistics were calculated for both the non-missing daily score data and the missing daily score data, by the number of days with missing data (ie, 4, 5, or 6 days). The data were compared using ratios.

The comparisons show that the means, medians, and standard deviations are almost the same for the non-missing and missing data, across the different combinations of overall missing rate and number of daily scores missing. Specifically, for the extreme case of 20% of patient clusters with 6 daily scores missing in each week, the weekly average change from baseline compared to the non-missing data shows that the means, medians, and standard deviations are almost the same. Given this finding, the results suggest that, for the current study, even if a patient has only 1 valid daily score within a week, it still can be used to represent the average for the week, to generate a robust analysis result. The summary is based on a total of 2000 replicated datasets, and for each of the replicated datasets the number of patients is 550.

Table 11: Simulation Results for WHSS During the 8-week Treatment Period

Week	Non-Missing			With Missing				Ratio (With Missing/Non-missing)		
				10% Missing in Each Replicated Data						
	Mean	Median	SD	Missing[a]	Mean	Median	SD	Mean	median	SD
1	-0.69	-0.7	1.04	4	-0.69	-0.7	1.05	1	1	1.01
2	-1.14	-1.1	1.13		-1.14	-1.1	1.14	1	1	1.01
3	-1.36	-1.3	1.18		-1.36	-1.3	1.19	1	1	1.01
4	-1.62	-1.7	1.24		-1.62	-1.7	1.25	1	1	1.01
5	-1.75	-1.7	1.30		-1.75	-1.7	1.31	1	1	1.01
6	-1.84	-1.9	1.22		-1.84	-1.9	1.23	1	1	1.01
7	-1.84	-2.0	1.29		-1.84	-2.0	1.30	1	1	1.01
8	-1.91	-1.9	1.34		-1.91	-1.9	1.35	1	1	1.01
1	-0.69	-0.7	1.04	5	-0.69	-0.7	1.07	1	1	1.03
2	-1.14	-1.1	1.13		-1.14	-1.1	1.15	1	1	1.02
3	-1.36	-1.3	1.18		-1.36	-1.3	1.20	1	1	1.02
4	-1.62	-1.7	1.24		-1.62	-1.7	1.26	1	1	1.02
5	-1.75	-1.7	1.30		-1.75	-1.7	1.32	1	1	1.01
6	-1.84	-1.9	1.22		-1.84	-1.9	1.24	1	1	1.02
7	-1.84	-2.0	1.29		-1.84	-2.0	1.31	1	1	1.01
8	-1.91	-1.9	1.34		-1.91	-1.9	1.36	1	1	1.01
1	-0.69	-0.7	1.04	6	-0.69	-0.7	1.10	1	1	1.06
2	-1.14	-1.1	1.13		-1.14	-1.1	1.18	1	1	1.05
3	-1.36	-1.3	1.18		-1.36	-1.3	1.23	1	1	1.04
4	-1.62	-1.7	1.24		-1.62	-1.7	1.28	1	1	1.04
5	-1.75	-1.7	1.30		-1.75	-1.7	1.34	1	1	1.03
6	-1.84	-1.9	1.22		-1.84	-1.9	1.26	1	1	1.04
7	-1.84	-2.0	1.29		-1.84	-2.0	1.33	1	1	1.03
8	-1.91	-1.9	1.34		-1.91	-2.0	1.38	1	1	1.03
				20% Missing in Each Replicated Data						
1	-0.69	-0.7	1.04	4	-0.69	-0.7	1.07	1	1	1.03
2	-1.14	-1.1	1.13		-1.14	-1.1	1.16	1	1	1.02
3	-1.36	-1.3	1.18		-1.36	-1.3	1.20	1	1	1.02
4	-1.62	-1.7	1.24		-1.62	-1.6	1.26	1	1	1.02
5	-1.75	-1.7	1.30		-1.75	-1.7	1.32	1	1	1.01
6	-1.84	-1.9	1.22		-1.84	-1.9	1.24	1	1	1.02
7	-1.84	-2.0	1.29		-1.84	-2.0	1.31	1	1	1.01
8	-1.91	-1.9	1.34		-1.91	-1.9	1.36	1	1	1.01
1	-0.69	-0.7	1.04	5	-0.69	-0.6	1.09	1	0.9	1.05
2	-1.14	-1.1	1.13		-1.14	-1.1	1.18	1	0.9	1.04

Week	Non-Missing			With Missing				Ratio (With Missing/Non-missing)		
				10% Missing in Each Replicated Data						
	Mean	Median	SD	Missing[a]	Mean	Median	SD	Mean	median	SD
3	-1.36	-1.3	1.18		-1.36	-1.3	1.22	1	1	1.04
4	-1.62	-1.7	1.24		-1.62	-1.6	1.28	1	1	1.03
5	-1.75	-1.7	1.30		-1.75	-1.7	1.34	1	1	1.03
6	-1.84	-1.9	1.22		-1.84	-1.9	1.25	1	1	1.03
7	-1.84	-2.0	1.29		-1.84	-2.0	1.33	1	1	1.03
8	-1.91	-1.9	1.34		-1.91	-2.0	1.38	1	1	1.02
1	-0.69	-0.7	1.04	6	-0.69	-0.7	1.16	1	1	1.12
2	-1.14	-1.1	1.13		-1.14	-1.1	1.23	1	0.9	1.09
3	-1.36	-1.3	1.18		-1.36	-1.3	1.28	1	1	1.08
4	-1.62	-1.7	1.24		-1.62	-1.7	1.33	1	1	1.07
5	-1.75	-1.7	1.30		-1.75	-1.7	1.38	1	1	1.06
6	-1.84	-1.9	1.22		-1.84	-2.0	1.30	1	1	1.07
7	-1.84	-2.0	1.29		-1.84	-2.0	1.37	1	1	1.06
8	-1.91	-1.9	1.34		-1.91	-2.0	1.42	1	1	1.06

a. Number of daily scores missing.

Table 12 shows summary results for the WRFS. These results show a similar pattern to that observed for WHSS and a similar conclusion can be drawn. That is, for the current study setting, the WRFS data are robust, and 1 daily score can be used to represent a weekly average value for data analysis. The summary is based on a total of 2000 replicated datasets, and for each of the replicated datasets the number of patients is 550.

Table 12: Simulation Results for WRFS During the 8-week Treatment Period

Week	Non-Missing			With Missing				Ratio (With Missing/Non-missing)		
				10% Missing in Each Replicated Data						
	Mean	Median	SD	Missing[a]	Mean	Median	SD	Mean	median	SD
1	-0.43	-0.3	0.76	4	-0.43	-0.3	0.77	1	1	1.02
2	-0.78	-0.7	0.93		-0.78	-0.7	0.94	1	1	1.01
3	-0.78	-0.7	1.02		-0.78	-0.7	1.03	1	1	1.01
4	-0.96	-0.8	1.06		-0.96	-0.8	1.06	1	1	1.01
5	-0.98	-0.8	1.06		-0.97	-0.8	1.07	1	1	1.01
6	-0.95	-0.9	1.00		-0.95	-0.9	1.01	1	1	1.01
7	-0.96	-0.9	1.04		-0.96	-0.9	1.05	1	1	1.01
8	-1.07	-1.1	1.11		-1.07	-1.1	1.12	1	1	1.01
1	-0.43	-0.3	0.76	5	-0.43	-0.3	0.78	1	1	1.03
2	-0.78	-0.7	0.93		-0.78	-0.7	0.95	1	1	1.02
3	-0.78	-0.7	1.02		-0.78	-0.7	1.04	1	1	1.02
4	-0.96	-0.8	1.06		-0.96	-0.8	1.07	1	1	1.01
5	-0.98	-0.8	1.06		-0.97	-0.8	1.07	1	1	1.01

Week	Non-Missing			With Missing				Ratio (With Missing/Non-missing)		
				10% Missing in Each Replicated Data						
	Mean	Median	SD	Missing[a]	Mean	Median	SD	Mean	median	SD
6	-0.95	-0.9	1.00		-0.95	-0.9	1.02	1	1	1.01
7	-0.96	-0.9	1.04		-0.96	-0.9	1.06	1	1	1.01
8	-1.07	-1.1	1.11		-1.07	-1.1	1.13	1	1	1.01
1	-0.43	-0.3	0.76	6	-0.43	-0.3	0.81	1	1	1.07
2	-0.78	-0.7	0.93		-0.78	-0.7	0.97	1	1	1.04
3	-0.78	-0.7	1.02		-0.78	-0.7	1.06	1	1	1.04
4	-0.96	-0.8	1.06		-0.96	-0.8	1.09	1	1	1.03
5	-0.98	-0.8	1.06		-0.98	-0.8	1.09	1	1.1	1.03
6	-0.95	-0.9	1.00		-0.95	-0.9	1.04	1	1	1.03
7	-0.96	-0.9	1.04		-0.96	-0.9	1.07	1	1	1.03
8	-1.07	-1.1	1.11		-1.07	-1.1	1.14	1	1	1.03
				20% Missing in Each Replicated Data						
1	-0.43	-0.3	0.76	4	-0.43	-0.3	0.78	1	1.1	1.03
2	-0.78	-0.7	0.93		-0.78	-0.7	0.95	1	1	1.02
3	-0.78	-0.7	1.02		-0.78	-0.7	1.04	1	1	1.02
4	-0.96	-0.8	1.06		-0.96	-0.8	1.07	1	1	1.01
5	-0.98	-0.8	1.06		-0.98	-0.8	1.07	1	1	1.01
6	-0.95	-0.9	1.00		-0.95	-0.9	1.02	1	1	1.02
7	-0.96	-0.9	1.04		-0.96	-0.9	1.06	1	1	1.01
8	-1.07	-1.1	1.11		-1.07	-1.1	1.13	1	1	1.01
1	-0.43	-0.3	0.76	5	-0.43	-0.3	0.81	1	1.1	1.06
2	-0.78	-0.7	0.93		-0.78	-0.7	0.97	1	1	1.03
3	-0.78	-0.7	1.02		-0.78	-0.7	1.05	1	1	1.03
4	-0.96	-0.8	1.06		-0.96	-0.8	1.08	1	1	1.03
5	-0.98	-0.8	1.06		-0.97	-0.8	1.09	1	1.1	1.03
6	-0.95	-0.9	1.00		-0.95	-0.9	1.03	1	1	1.03
7	-0.96	-0.9	1.04		-0.96	-0.9	1.07	1	1	1.03
8	-1.07	-1.1	1.11		-1.07	-1.1	1.14	1	1	1.02
1	-0.43	-0.3	0.76	6	-0.43	-0.3	0.87	1	0.9	1.14
2	-0.78	-0.7	0.93		-0.78	-0.7	1.01	1	1	1.08
3	-0.78	-0.7	1.02		-0.78	-0.7	1.10	1	1	1.07
4	-0.96	-0.8	1.06		-0.96	-0.9	1.12	1	1.1	1.06
5	-0.98	-0.8	1.06		-0.97	-0.9	1.13	1	1.1	1.06
6	-0.95	-0.9	1.00		-0.95	-0.9	1.07	1	1.1	1.07
7	-0.96	-0.9	1.04		-0.96	-1.0	1.11	1	1	1.06
8	-1.07	-1.1	1.11		-1.07	-1.1	1.18	1	0.9	1.06

a. Number of daily scores missing.

Proportion of Heartburn-free Days During the 8-week Treatment Period

A similar bootstrapping resampling process for patient clusters that was used for the parameters WHSS and WRFS can be applied to PHFD. The descriptive summary statistics were calculated for both the non-missing daily score data and the missing daily score data, by the number of days with missing data (ie, 4, 5, or 6 days). The results are presented in Table 13.

The comparisons show that the means, medians, and standard deviations are almost the same for the non-missing and missing data, across different combinations of overall missing rate and number of daily scores missing. This suggests that the algorithm for calculating the PHFD (the number of heartburn-free days divided by the number of diary entry days) is robust. The summary is based on a total of 2000 replicated datasets, and for each of the replicated datasets the number of patients is 550.

Table 13: Simulation Results for PHFD During the 8-week Treatment Period

Non-Missing			With Missing				Ratio (With Missing/Non-missing)		
			10% Missing in Each Replicated Data						
Mean	Median	SD	Missing[a]	Mean	Median	SD	Mean	median	SD
0.307	0.2	0.315	4	0.307	0.2	0.315	1	1	1
0.307	0.2	0.315	5	0.308	0.2	0.315	1	1	1
0.307	0.2	0.315	6	0.307	0.2	0.316	1	1	1
			20% Missing in Each Replicated Data						
0.307	0.2	0.315	4	0.307	0.2	0.315	1	1	1
0.307	0.2	0.315	5	0.308	0.2	0.316	1	1	1
0.307	0.2	0.315	6	0.308	0.2	0.317	1	1	1.01

a. Number of daily scores missing.

Appendix 2. Visit Window Slotting Conventions

Visit Time Windows for Safety Analysis

Table 14 below presents the visits assigned for the safety analysis corresponding to the range of trial days (window) during which an actual visit may have occurred.

Table 14: Visit Time Windows for Safety Analysis

Derived Visit	*Scheduled Test / Visit Day*[a]	*Window*
Baseline	Day 1	Days ≤1
Week 4 (Day 29) Visit	Day 29	Days [2, 43]
Week 8 (Day 57) Visit	Day 57	Days ≥44
End of Treatment[b]	End of Treatment Visit	

a. Relative to the day of randomization (Day 1)
b. "End of Treatment" will be presented in analysis tables for safety parameters, including ECG, clinical laboratory, and vital signs.

Test/Visit Day will be calculated as follows: test/visit date – date of randomization + 1.

If there are multiple values available for the same Visit, the data with the latest date for that visit will be utilized for summary.

Visit Time Windows for Efficacy Analysis

Table 15 presents the analysis weeks assigned for the efficacy analysis of the patient diary data related to daily GERD symptom assessment. These analysis weeks will be used in the calculations for all week-based parameters (eg, WHSS, WRFS, etc.).

Table 15: Analysis Time Windows for Efficacy Analysis – Daily Assessments

Period	**Analysis Week**	**Begins**[a]	**Ends**[a]
Pretreatment (Baseline[b])	Week -2	Day -14	Day -8
	Week -1	Day -7	Day -1 (Day before randomization) [c]
Treatment	Week 1	Day 1 (Day of Randomization) [d]	Day 7
	Week 2	Day 8	Day 14
	Week 3	Day 15	Day 21
	Week 4	Day 22	Day 28
	Week 5	Day 29	Day 35
	Week 6	Day 36	Day 42
	Week 7	Day 43	Day 49
	Week 8	Day 50	Day 56
	Week 9	Day 57	Day 63

a. Relative to the day of randomization (Day 1). There is no Day 0.
b. Baseline for efficacy parameters will be derived from the eDiary collected for Pretreatment Week -1 (Week -2 if data for Week -1 not available), specifically for the time dependent proportion parameters (defined as days with the event of interesting divided by the number of diary entry days during the relevant period) the last 14 days (Week -2 to Week -1) prior to day of randomization will be used.

c. Day -1 for efficacy parameters derived from the morning diary.
d. Day 1 for efficacy parameters derived from the morning diary.
e. All values that were assessed prior to or on the day of the last dose will be used for the efficacy analysis. However, all values, regardless of whether they were collected prior to or after the last dose, will be used for the sensitivity analysis.

Table 16 presents the analysis weeks assigned for the efficacy analysis of the patient diary data related to Weekly Questions (eg, degree of relief questions, symptom bothersomeness questions, and treatment satisfaction questions).

Table 16: Analysis Time Windows for Efficacy Analysis – Weekly Assessments

Period	Analysis Week	Begins[a]	Ends[a]
Pretreatment (Baseline[b])	Week -2	Day -11	Day -5
	Week -1	Day -4	Day -1 (Day before randomization)
Treatment	Week 1	Day 7	Day 10
	Week 2	Day 11	Day 17
	Week 3	Day 18	Day 24
	Week 4	Day 25	Day 31
	Week 5	Day 32	Day 38
	Week 6	Day 39	Day 45
	Week 7	Day 46	Day 52
	Week 8	Day 53	Day 59
	Week 9	Day 60	Day 66

a. Relative to the day of randomization (Day 1). There is no Day 0.
b. Baseline values for efficacy parameters will be derived from the eDiary collected in the Pretreatment Period Week -1 (Week -2 if data from Week -1 not available), unless otherwise specified.
c. All values that were assessed prior to or on the day of the last dose will be used for the efficacy analysis. However, all values, regardless of whether they were collected prior to or after the last dose, will be used for the sensitivity analysis.

In general, weekly questions will be assigned only to the analysis week for which the question covers at least 4 days of that week.

If patients answer the weekly questions multiple times for a week, for each question, the average of the respective answers will be assigned as the weekly value.

Appendix 3. Pooling of Trial Centers

Because of the potential of many study centers to have a small number of patients, the centers will be pooled by the following 6 geographic regions (as listed in Table 17): US Northeast, US Southeast, US Midwest, US Southwest, US West and Canada.

Table 17: Definition of Geographic Regions

US Northeast	US Southeast	US Midwest	US Southwest	US West	Canada
CT	AL	IA	AZ	CA	CANADA
DE	AR	IL	NM	CO	-
MA	FL	IN	OK	ID	-
MD	GA	KS	TX	MT	-
ME	KY	MI	-	NV	-
NH	LA	MN	-	OR	-
NJ	MS	MO	-	UT	-
NY	NC	ND	-	WA	-
PA	SC	NE	-	WY	-
RI	TN	OH	-	-	-
VT	VA	SD	-	-	-
-	WV	WI	-	-	-

Appendix 4. Schedule of Evaluations

Table 18: Schedule of Evaluations

	Screening Period (Up to 5 weeks)	**Pretreatment Period (2 weeks)**	**Treatment Period (8 weeks)**			**Follow-up**
Visit Days →	Screening Visit (Day -56 to Day -15)	Pretreatment Visit (Day -21 to Day -1)	Randomization Visit (Day 1)	Week 4 Visit (Day 29 ± 3)	Week 8 / End-of-Treatment Visit (Day 57 + 3)	Safety Phone Call (EOT + 7d)
Visit Numbers →	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	
Study Procedure ↓						
Inclusion and Exclusion Criteria Verification	X	X	X			
Signing of ICF	X					
IWRS Registration (a)	X	X	X	X	X	
Demographics	X					
Medical & Disease History	X					
Physical Examination (b)	X				X	
Body Weight and Height (c)	X	X	X	X	X	
H2RA/Antacid Washout (d)	X					
EGD (e)	X				X	
Up to 96 Hours of pH Testing with Bravo Device (f)	X					
Seated Vital Signs (g)	X		X	X	X	
12-Lead ECG (h)	X		X		X	
Prior and Concomitant Medications & Procedures (i)	X	X	X	X	X	X
Clinical Laboratory Tests (j)	X		X	X	X	
Pregnancy Test (k)	X		X	X	X	
Drug and Alcohol Screen (l)	X					
AE Evaluations (m)	X	X	X	X	X	X
Rescue Medicine Dispensed (n)		X	X	X		
eDiary Training and Dispensation		X				
eDiary (o)		X	X	X	X	

	Screening Period (Up to 5 weeks)	**Pretreatment Period (2 weeks)**	**Treatment Period (8 weeks)**			**Follow-up**
Visit Days →	Screening Visit (Day -56 to Day -15)	Pretreatment Visit (Day -21 to Day -1)	Randomization Visit (Day 1)	Week 4 Visit (Day 29 ± 3)	Week 8 / End-of-Treatment Visit (Day 57 + 3)	Safety Phone Call (EOT + 7d)
Visit Numbers →	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	
Study Procedure ↓						
Weekly Symptom and Treatment Assessments		X	X	X	X	
WPAI-Sleep-GERD			X	X	X	
EQ-5D-3L			X	X	X	
Randomization			X			
Study Medication Dispensed (p)			X	X		
Study Medication Return (q)				X	X	
eDiary Return					X	
Follow-up Phone Call (r)						X

Abbreviations: AE = adverse event; BP = blood pressure; CBC = complete blood count; ECG = electrocardiogram; EGD = esophagogastroduodenoscopy; EOT = End-of-treatment; EQ = EuroQol; H2RA = histamine-2 receptor antagonist; HEENT = head, eyes, ears, nose, and throat; ICF = informed consent form; IWRS = interactive web response system; PPI = proton pump inhibitor; mRESQ-eD = modified Reflux Symptom Questionnaire Electronic Diary; SAE = serious adverse event

a. Site personnel will interact with IWRS to register the patient visit and transition the patient to the next appropriate study period.
b. A physical examination should include the following assessments: general appearance; HEENT; neck; cardiovascular; respiratory; abdomen/liver/spleen; musculoskeletal; lymph nodes; skin; neurologic; and mental status. Breast, genitourinary, and rectal examinations are optional and may be performed at the discretion of the Investigator.
c. Height will be measured only at the Screening Visit.
d. During the Screening Period, patients will washout H2RAs at least 5 days prior, and antacids 1 day prior, to the EGD and Bravo testing. Patients may resume antacid use upon completion of the Bravo testing, but must refrain from H2RA use for the remainder of the study. During the Pretreatment Period, patients will refrain from using any anti-reflux medications, antacids, and H2RAs, except for the liquid antacid that is provided as rescue medication (aluminum hydroxide / magnesium hydroxide).
e. All patients will be required to undergo an EGD during the Screening Period. There should be a minimum of 7 days between the EGD and the start of the Pretreatment Period to allow for pH testing and patient stabilization. A repeat EGD will be performed at the Week 8 / EOT Visit for all patients who have completed at least 4 weeks of treatment and had erosive esophagitis on the screening EGD (based on the Los Angeles [LA] classification of esophagitis, see Appendix 8) as determined by either the site personnel or the central reader.
f. Up to 96 hours of pH testing with the Bravo device. If for some reason 96 hours of testing is not possible, then approximately 48 hours of testing is acceptable.

g. Vital sign measurements include oral temperature (°C), respiratory rate, systolic and diastolic BP, and pulse. Respiratory rate, BP, and pulse measurements must be obtained after the patient has been seated for at least 5 minutes.
h. 12-Lead ECGs should be obtained after the patient has been supine for at least 5 minutes.
i. Prior medications will be collected at the Screening Visit as follows: all medicines taken by the patient during the 30 days before the Screening Visit, most recent use of PPIs, H2RAs, and antacids. Concomitant medications and procedures will be collected in the EDC system.
j. Clinical laboratory tests include clinical chemistry, hematology (CBC), coagulation, and urinalysis. If the triglyceride value exceeds the protocol-specified criteria and the patient was not under fasted conditions, the patient may return to complete a fasted lipid panel.
k. For all female patients of childbearing potential, a negative urine or serum pregnancy test must be performed and documented at the Screening Visit, and at the Randomization Visit (before dosing) in order for the patient to be randomized into the study. A urine or serum pregnancy test will also be obtained at the Week 4 and EOT Visits. All positive urine pregnancy test results will be confirmed by a serum pregnancy test.
l. Patients must undergo a urine drug screen for selected drugs of abuse (cocaine, barbiturates, amphetamines, opiates, benzodiazepines, and cannabinoids) and a serum alcohol screen at the Screening Visit.
m. For randomized patients, all AEs will be captured from the time the patient signs the ICF through the Follow-up Phone call.
n. Rescue medicine (aluminum hydroxide / magnesium hydroxide) will be supplied to patients at the Pretreatment Visit, and if needed, at subsequent visits.
o. The eDiary will be dispensed at the Pretreatment Visit. Patients must complete daily questions at least 5 days each week during the 14 days before the Treatment Period and must complete the weekly questions at least once during the 7 days before the Treatment Period in order to be eligible for randomization. Patients should bring their eDiary to each visit. The eDiary will be used for daily assessments (PPI administration, mRESQ-eD, dyspepsia symptoms, sleep disturbance, and rescue medication use), and weekly assessments (degree of relief, symptom bothersomeness, symptom relief, treatment satisfaction, and questionnaires).
p. The first dose of study medication will be administered in the clinic with liquid and a snack at the Randomization Visit. At all other visits, patients will take study medication prior to arriving at the clinic but will be dispensed additional doses needed until the next study visit.
q. Treatment adherence with study drug will be assessed based on return of unused tablets.
r. Study site will contact each patient via telephone 7 days after the EOT Visit to collect information pertaining to ongoing AEs/SAEs/concomitant medications/concomitant therapies, and information concerning any new AEs/SAEs /concomitant medications/concomitant therapies since the EOT Visit.

Appendix 5. Modified Reflux Symptom Questionnaire Electronic Diary

Instructions: Please answer the following questions to help us understand the symptoms you experienced over the past 24 hours because of your reflux disease. For each question, please choose the answer most appropriate for you.

1.	Over the past 24 hours, how would you rate the severity of your burning feeling behind the breastbone or in the center of the upper stomach? *0=Did not have/ 1=Very mild/ 2=Mild/ 3=Moderate/ 4=Moderately severe/ 5=Severe*
2.	Over the past 24 hours, how would you rate the severity of your pain behind the breastbone or in the center of the upper stomach? *0=Did not have/ 1=Very mild/ 2=Mild/ 3=Moderate/ 4=Moderately severe/ 5=Severe*
3.	Over the past 24 hours, how would you rate the severity of your difficulty swallowing? *0=Did not have/ 1=Very mild/ 2=Mild/ 3=Moderate/ 4=Moderately severe/ 5=Severe*
4.	Over the past 24 hours, how would you rate the severity of your hoarseness? *0=Did not have/ 1=Very mild/ 2=Mild/ 3=Moderate/ 4=Moderately severe/ 5=Severe*
5.	Over the past 24 hours, how would you rate the severity of your cough? *0=Did not have/ 1=Very mild/ 2=Mild/ 3=Moderate/ 4=Moderately severe/ 5=Severe*
6.	Over the past 24 hours, how often did you experience regurgitation (liquid or food moving upwards towards your throat or mouth)? *0=Never/ 1=Rarely/ 2=Sometimes/ 3=Often/ 4=Very often*
7.	Over the past 24 hours, how often did you experience an acid or bitter taste in the mouth? *0=Never/ 1-Rarely/ 2=Sometimes/ 3=Often/ 4=Very often*
8.	Over the past 24 hours, how often did you experience burping? *0=Never/ 1=Rarely/ 2=Sometimes/ 3=Often/ 4=Very often*

Appendix 6. Other Daily Assessments

Heartburn

The following question will be answered by the patient in the evening diary.

Over the past 24 hours, how would you rate the severity of your heartburn?

0=Did Not Have, 1=Very Mild, 2=Mild, 3=Moderate, 4=Moderately Severe, 5=Severe

Sleep Disturbance Due to GERD

The following questions will be answered by the patient upon getting up each morning (5:00 a.m. to 12:00 p.m.) using the eDiary:

Falling Asleep

Last night, did you have trouble falling asleep because of GERD symptoms?

[Yes, No]

Awakenings

Last night, how many times did you wake up during the night because of GERD symptoms?

0

1 time

2 times

3-5 times

6 or more times

Use of per-protocol rescue medicine (antacid)

How many times did you use your rescue medicine (liquid antacid) during the past 24 hours?

0

1 time

2 times

3-5 times

6 or more times

Appendix 7. Weekly Assessments

Degree of Relief Assessments

Administered weekly during Pretreatment and Treatment Periods:

How would you rate your heartburn (a pain or burning sensation in your chest, behind the breastbone) over the past 7 days?

How would you rate your regurgitation (the feeling of stomach contents, either liquid or food, moving upwards to your throat or mouth) over the past 7 days?

How would you rate your overall GERD symptoms over the past 7 days?

Response Scale for all Degree of Relief Assessments:

1=Significantly relieved

2=Moderately relieved

3=Somewhat relieved

4=Unchanged

5=Somewhat worse

6=Moderately worse

7=Significantly worse

Bothersomeness Assessments

Administered weekly during Pretreatment and Treatment Periods:

How much were you bothered by heartburn (a pain or burning sensation in your chest, behind the breastbone) over the past 7 days?

How much were you bothered by regurgitation (the feeling of stomach contents, either liquid or food, moving upwards to your throat or mouth) over the past 7 days?

How much were you bothered by your overall GERD symptoms over the past 7 days?

1=Not at all

2=A little bit

3=A moderate amount

4=A great deal

5=An extreme amount

Treatment Satisfaction Assessment

Administered weekly during Treatment Period:

How would you rate your satisfaction with the study treatment?

1=Very dissatisfied

2=Dissatisfied

3=Neither satisfied nor dissatisfied

4=Satisfied

5=Very satisfied

Appendix 8. Los Angeles Classification of Esophagitis

Classification	Definition
Not Present	No breaks (erosions) in the esophageal mucosa. (However, edema, erythema, or friability may be present.)
Grade A	One (or more) mucosal break no longer than 5 mm that does not extend between the tops of two mucosal folds.
Grade B	One (or more) mucosal break more than 5 mm long that does not extend between the tops of two mucosal folds.
Grade C	One (or more) mucosal break that is continuous between the tops of two or more mucosal folds but which involve less than 75% of the circumference.
Grade D	One (or more) mucosal break which involves at least 75% of the esophageal circumference.