Title: A Randomized, Double-Blind, Proof-of-Concept, Phase 2 Study to Evaluate the Efficacy and Safety of Once Daily Oral Vonoprazan 20 mg or Vonoprazan 40 mg Compared to Esomeprazole 40 mg for the Treatment of Subjects With Symptomatic Gastro-Esophageal Reflux Disease Who have a Partial Response Following Treatment with a High Dose of Proton Pump Inhibitor

NCT Number: NCT02743949

SAP Approve Date: 15 January 2019

Certain information within this Statistical Analysis Plan has been redacted (ie, specific content is masked irreversibly from view with a black/blue bar) to protect either personally identifiable (PPD) information or company confidential information (CCI).

This may include, but is not limited to, redaction of the following:

- Named persons or organizations associated with the study.
- Proprietary information, such as scales or coding systems, which are considered confidential information under prior agreements with license holder.
- Other information as needed to protect confidentiality of Takeda or partners, personal information, or to otherwise protect the integrity of the clinical study.

TAKEDA DEVELOPMENT CENTER

STATISTICAL ANALYSIS PLAN

STUDY NUMBER: VONOPRAZAN-2001

A Randomized, Double-Blind, Proof-of-Concept, Phase 2 Study to Evaluate the Efficacy and Safety of Once Daily Oral Vonoprazan 20 mg or Vonoprazan 40 mg Compared to Esomeprazole 40 mg for the Treatment of Subjects With Symptomatic Gastro-Esophageal Reflux Disease Who have a Partial Response Following Treatment with a High Dose of Proton Pump Inhibitor

PHASE 2

Version**: 1**

Date**:** 15 January 2019

Prepared by:

PPD

1.0 APPROVAL SIGNATURES

Electronic signatures can be found on the last page of this document.

Study Title: A Randomized, Double-Blind, Proof-of-Concept, Phase 2 Study to Evaluate the Efficacy and Safety of Once Daily Oral Vonoprazan 20 mg or Vonoprazan 40 mg Compared to Esomeprazole 40 mg for the Treatment of Subjects With Symptomatic Gastro-Esophageal Reflux Disease Who have a Partial Response Following Treatment with a High Dose of Proton Pump Inhibitor

TDC Approvals:

PPD

2.0 TABLE OF CONTENTS

3.0 LIST OF ABBREVIATIONS

AE	adverse event
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
ANOVA	analysis of variance
AST	aspartate aminotransferase
AUC_{τ}	area under the curve from time zero to tau
BMI	body mass index
BUN	blood urea nitrogen
CI	confidence interval
C_{max}	maximum observed plasma concentration
CPK	creatine phosphokinase
eCRF	electronic case report form
ECG	electrocardiogram
EM	extensive metabolizer
EOT	end of treatment
EQ-5D-5L	EuroQol-5 dimension-5 levels
ESS	enrolled subjects set
FAS	full analysis set
GERD	gastro-esophageal reflux disease
GGT	γ-glutamyl transferase
GI	gastrointestinal
HADS	hospital depression and anxiety scale
HIV	human immunodeficiency virus
H pylori	helicobacter pylori
HRQoL	health related quality of life
ICH	International Conference on Harmonisation
IVRS	interactive voice response system
LA	Los Angeles
LDH	lactate dehydrogenase
LLN	lower limit of normal
LOCF	last observation carried forward
MAV	markedly abnormal values
MedDRA	Medical Dictionary for Regulatory Activities
PAGI-QOL	Patient Assessment of Upper Gastrointestinal Disorders Quality-of-Life Index
PAGI-SYM	Patient Assessment of Upper Gastrointestinal Disorders Symptom Severity Index
PD	pharmacodynamics
PGIC	patient global impression of change
PK	pharmacokinetics

PM	poor metabolizer
PPI	proton pump inhibitor
PPS	per protocol set
PRO	patient-reported outcome
PSQI	Pittsburgh sleep quality index
PT	preferred term
PTE	pre-treatment event
QD	quaque die (once a day)
QOL	quality-of-life
QTcF	corrected QT using the Fredericia correction
RESQ-eD	reflux symptom questionnaire e-Diary
SAE	serious adverse event
SAF	safety analysis set
SAP	statistical analysis plan
SD	standard deviation
SDB	standard database
SE	standard error
SOC	system organ class
TAU	therapeutic area unit
TEAE	treatment emergent adverse event
TLGs	tables, listings, and graphs
t_{max}	time to C_{max}
UGI	upper gastrointestinal
ULN	upper limit of normal
VAS	visual analogue scale
WHODrug	World Health Organization Drug Dictionary

4.0 OBJECTIVES

4.1 PRIMARY OBJECTIVES

To determine the effect of vonoprazan compared to esomeprazole for preventing heartburn symptoms over a 4-week treatment period in subjects who have a partial response to treatment with esomeprazole.

4.2 SECONDARY OBJECTIVES

To determine the effect of vonoprazan treatment on sustained resolution of heartburn symptoms over a 4 week treatment period (at least one 7-day symptom-free period).

4.3 ADDITIONAL OBJECTIVES

To determine the effect of vonoprazan compared with esomeprazole on GERD-related endpoints, including:

- Preventing heartburn at night over a 4 week treatment period.
- Sustained resolution of heartburn during the final week of treatment.
- Preventing regurgitation during 24-hour periods, over a 4 week treatment period.
- Preventing regurgitation at night over a 4 week treatment period.
- Preventing heartburn and/or regurgitation during 24-hour periods over a 4 week treatment period.
- Complete prevention of regurgitation (response rate) during the final week of treatment.
- Evaluation of the effect of vonoprazan on the use of rescue medication over 4 weeks of treatment.
- Evaluation of the effect of vonoprazan on GERD quality of sleep over 4 weeks of treatment.
- Evaluation of the effect of vonoprazan on Health Related Quality of Life (HRQoL) over 4 weeks of treatment.

To evaluate the pharmacokinetics of vonoprazan in patients.

To determine the safety and tolerability of vonoprazan in this population.

To evaluate the pharmacokinetics and clinical efficacy of vonoprazan in patients identified as CYP2C19 poor or extensive metabolizers (PM/EM).

In this study, samples for pharmacogenomics will be collected where possible and stored for possible exploratory investigation of drug response or disease. In this study, using Vonoprazan, or in a set of clinical trials, if variability is seen in responsiveness to study medication and it is

suspected to be attributable to gene polymorphism, pharmacogenomics analyses may be conducted to explore gene polymorphism relationships, as indicated by the observations.

4.4 STUDY DESIGN

This is a phase 2, randomized, double-blind, parallel-group, 3-arm Proof of Concept study with a 4-week active Treatment Period. This study will evaluate the safety and efficacy of vonoprazan 20 mg once daily (QD) and 40 mg QD compared with esomeprazole 40 mg QD when administered as an oral daily dose in the morning, at least 1 hour before the first meal of the day, for 4 weeks of treatment in subjects with symptomatic GERD who have a partial response to treatment with a high dose of esomeprazole. This study will be conducted at a total of up to 50 sites in Europe. A total of 213 subjects will be randomized in a 1:1:1 ratio of Vonoprazan 40 mg to Vonoprazan 20 mg to Esomeprazole 40 mg (71 per treatment arm).

Subjects will be eligible for participation in the study if they: have a documented history of symptoms of both heartburn and acid regurgitation prior to entry into the study; have a history of persistent heartburn and/or regurgitation symptoms that are troublesome despite an adequate course of proton pump inhibitor (PPI) treatment; continue to have symptoms of heartburn (and/or regurgitation) following 4 weeks treatment with a high dose of esomeprazole 40 mg QD during the Run-in Period; have symptoms of heartburn which increases following a 2-week Washout Period in the presence of regurgitation prior to randomization; and meet all of the inclusion criteria and none of the exclusion criteria. Patients with mild (LA grade A) esophagitis are also permitted to enter the study. To confirm LA grade, endoscopy will be performed at screening, or historical data of endoscopy up to one year prior to screening may be permitted. The study will consist of 3 periods: a 7-week Run-in period (initial 1 week screening period and a 6-week single-blind Run-in), a 4-week double blind Treatment Period, and a 1-week Safety Follow-up. The Run-in includes a 1-week general Screening Period during which the subject will remain on their prescribed PPI (Day-49 to Day -42); followed by a 4-week single-blind PPI assessment period with blinded esomeprazole 40 mg QD (Day -42 to Day -14); and then a 2-week single-blind placebo Off-PPI Assessment period (Day -14 to Day-1). Subjects who remain symptomatic and compliant will be randomized to the 4-week Treatment Period followed by a one week safety Follow-up Call. Certain medications will require washout prior to entry into the study and will not be permitted during the study.

The Screening Period will take up to 49 days. After providing informed consent and meeting the appropriate study entry criteria during the Initial Screening Visit (Visit 1), subjects will be taken off their prescribed PPI and be given esomeprazole 40 mg QD, be provided with an electronic diary (Visit 2), and trained in how to complete it. They will return 4 weeks later to assess treatment compliance, completion of the e-Diary and assessment of the PPI Treatment Period (Day -21 to -14). If the subject is still symptomatic, he or she will receive single-blinded placebo medication for a 2-week PPI Washout Period and their Off-PPI Baseline will be assessed during the week prior to randomization (Day -8 to -1).

It is estimated that approximately 426 subjects will be screened to ensure that a total of 213 subjects who meet the symptomatic GERD entry criteria for a partial response to PPI (defined as having heartburn following 4 weeks of treatment with esomeprazole 40 mg and an increase of at least 2 symptom days of heartburn (and/or regurgitation) in the last week of the Off-PPI assessment period compared with the last week of esomeprazole treatment) are randomized equally to 1 of 3 treatments at Visit 4 (vonoprazan 20 mg QD, vonoprazan 40 mg QD or esomeprazole 40 mg QD for 4 weeks). A 50% screen failure rate is higher than for standard GERD studies and has been estimated based on the duration of the run-in, inclusion of a 2 week placebo washout period and the symptom criteria to be met prior to randomization. Patients will take their first dose of study drug on the morning after Visit 4 and be treated on an outpatient basis, complete their e-Diary twice daily and will return to the clinic at Weeks 2 (Visit 5) and Week 4 (Visit 6) of the Treatment Period.

A Safety Follow-up call will be made approximately 1 week after completion of the 4-week Treatment Period. If possible, subjects who prematurely withdraw from the study will be seen for a withdrawal visit as soon as possible after discontinuation, and also will be contacted for Safety Follow-up approximately 1 week after withdrawal.

Rescue antacid medication will be provided for all subjects at the Initial Screening Visit and as required throughout the study.

From the Initial Screening Visit to the End-of-Treatment Visit, there will be 6 subject visits and a phone call during the Follow-up period. A schematic of the study design is included as Figure 1.

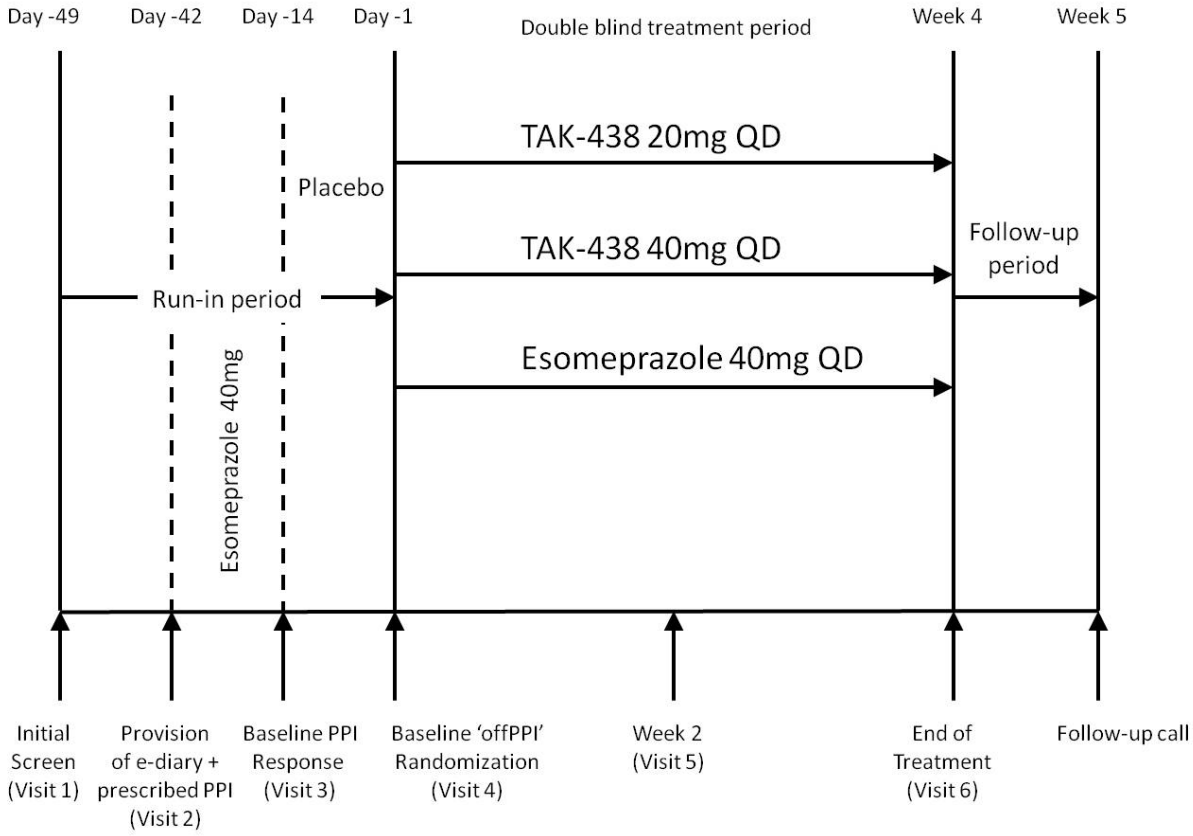

Figure 1 Schematic of Study Design

5.0 ANALYSIS ENDPOINTS

Primary Endpoints

Percentage of heartburn-free 24 hour periods (day and night) during 4 weeks of randomized double-blind treatment.

Secondary Endpoints

Proportion of subjects with ≥1 sustained resolution of heartburn (defined as ≥7 consecutive days without both daytime and nighttime heartburn any time during the 4-week randomized double-blind Treatment Period).

Additional Endpoints

Heartburn Endpoints

Change from baseline in percentage of heartburn-free 24 hour periods (day and night) during 4 weeks of randomized double-blind treatment, where baseline is defined as the 4 weeks run-in period on esomeprazole.

Change from baseline in proportion of subjects with ≥1 sustained resolution of heartburn, where baseline is defined as the 4 weeks run-in period on esomeprazole

Percentage of heartburn-free nights during 4 weeks of randomized double-blind treatment.

Percentage of heartburn-free days after 4 weeks of randomized double-blind treatment.

Proportion of patients with no nighttime heartburn symptoms during the last week of treatment.

Proportion of patients with no daytime heartburn symptoms during the last week of treatment.

Proportion of patients with no daytime and nighttime heartburn symptoms during the last week of treatment.

Percentage of subjects with ≥80% heartburn free days and nights during 4 weeks of randomized double-blind treatment.

Percentage of subjects with ≥80% heartburn free nights during 4 weeks of randomized double-blind treatment.

Percentage of subjects with ≥80% heartburn free days during 4 weeks of randomized double-blind treatment.

Percentage of heartburn free 24 hour periods (day and night) during weeks 1, 2 and week 3 of randomized double-blind treatment.

Time to first sustained resolution (defined as 7 consecutive days without daytime and nighttime heartburn).

Proportion of subjects who achieve sustained resolution of heartburn (defined as ≥7 consecutive days without daytime and nighttime heartburn) with an onset within 3 days of starting randomized double-blind treatment.

Proportion of subjects with resolution of symptoms with an onset within 3 days of starting randomized double-blind treatment and continuing out to the end of week 2 and end of week 4.

Maintenance of heartburn-free period – duration of heartburn symptom-free days and nights (24-hour periods) following the first 24-hour period when no symptoms were recorded until the next time symptoms were recorded.

Heartburn severity (severity of the most intense episode will be assessed for daytime and nighttime).

Percentage of responders, where a responder is defined as having at least 3 more days of not more than mild heartburn symptoms on average per week during the whole double blind treatment period compared to baseline, where baseline is defined as the 1 week run-in period on placebo prior to randomization.

Regurgitation Endpoints

Percentage of regurgitation-free nights during 4 weeks of randomized double-blind treatment.

Percentage of regurgitation-free days after 4 weeks of randomized double-blind treatment.

Percentage of regurgitation-free 24 hour periods during 4 weeks of randomized double-blind treatment.

Percentage of subjects with ≥80% regurgitation-free days and nights during 4 weeks of randomized double-blind treatment.

Percentage of subjects free from regurgitation ("no regurgitation") during week 1, 2 and 3 of randomized double-blind treatment.

Time to first sustained resolution (defined as 7 consecutive days without daytime and nighttime regurgitation).

Proportion of subjects who achieve sustained resolution of regurgitation (defined as ≥7 consecutive days without daytime and nighttime heartburn) with an onset within 3 days of starting randomized double-blind treatment.

Proportion of subjects with resolution of regurgitation symptoms with an onset within 3 days of starting randomized double-blind treatment and continuing out to the end of week 2 and the end of week 4.

Proportion of patients with no daytime and nighttime regurgitation symptoms during the last week of treatment.

Maintenance of regurgitation-free period – duration of regurgitation symptom-free days and nights (24-hour periods) following the first 24-hour period when no symptoms were recorded.

Regurgitation severity (severity of the most intense episode and will be assessed for daytime and nighttime).

Heartburn and Regurgitation Endpoint

Percentage of heartburn and/or regurgitation-free 24 hour periods (day and night) during 4 weeks of randomized double-blind treatment.

Rescue medication Endpoint

Percentage of days without using rescue medication.

Percentage of nights without using rescue medication.

Percentage of days and nights without using rescue medication.

GERD Sleep quality measure

Change from baseline in measured scores.

Proportion of subjects with clinically significant improvement on sleep disturbance measures.

HRQoL

HRQoL will be assessed using patient reported outcome questionnaires at patient visit and the following endpoints will be used for analysis:

– Change from baseline in HRQoL scores.

– Proportion of subjects with clinically significant improvement on HRQoL measures.

Other Additional Endpoints

Summary statistics of plasma concentrations of Vonoprazan at the Week 2 and 4 visit and summary pharmacokinetic parameters where available: maximum observed plasma concentration (C_{max}), time to reach C_{max} (t_{max}), area under the plasma concentration-time curve from time 0 to tau (AUC_{τ}).

Summary statistics of plasma concentrations of Vonoprazan at the Week 2 and 4 visit for CYP2C19 poor or extensive metabolizers (PM/EM) and summary pharmacokinetic parameters where available: maximum observed plasma concentration (C_{max}), time to reach C_{max} (t_{max}), area under the plasma concentration-time curve from time 0 to tau (AUC_{τ}).

Summary statistics of plasma concentrations of esomeprazole at the Week 4 visit.

Safety parameters examined will include AEs, serious adverse events (SAEs), vital signs, results of standard laboratory tests.

6.0 DETERMINATION OF SAMPLE SIZE

It is estimated that approximately 426 subjects will be screened from which a total of 213 subjects (71 per treatment arm) who have a partial response to treatment with a high dose of esomeprazole (40 mg QD for 4 weeks) are planned to be randomized into the study with the expectation that 180 subjects will complete the study (assuming a 15% dropout rate). The sample size of 60 subjects per treatment group will provide at least 80% power at the 2-sided 0.025 level of significance to detect an absolute 20% difference between each vonoprazan dose and esomeprazole 40 mg for the percentage of heartburn free 24 hour periods (day and night) during 4 weeks of randomized double-blind treatment. The pooled standard deviation was estimated as 33.61 from prior Dexlansoprazole (T-GD04-082, T-GD04-083, T-EE04-084, T-EE04-085) and vonoprazan (CCT-201) studies (Data on file). The sample size was calculated using the Wilcoxon rank-sum test in EAST version 6.0.

With the expectation that 180 subjects will complete the study, the secondary endpoint; proportion of subjects with ≥1 sustained resolution of heartburn (defined as ≥7 consecutive days without both daytime and nighttime heartburn any time during the 4-week randomized double-blind Treatment Period) has 80% power at the 2-sided 0.025 level of significance to detect an absolute 30% difference between each vonoprazan dose and esomeprazole, assuming a response rate of 33% for esomeprazole. The sample size was calculated using the Fisher's exact test in EAST version 6.0.

7.0 METHODS OF ANALYSIS AND PRESENTATION

7.1 GENERAL PRINCIPLES

There are four baselines defined for the study.

- Change in the primary and secondary endpoint from baseline, where baseline is defined as the 4 weeks run-in period on esomeprazole.
- Change in the PRO questionnaire scores from baseline, where baseline is the score at visit 3 ("baseline PPI response").
- Change in the PRO questionnaire scores from baseline, where baseline is the score at visit 4 ("baseline off-PPI response").
- Change in ECG, vital signs, laboratory results, physical examination, where baseline is defined as the last assessment prior to first dose of study drug on Day 1 (visit 4).

Note that in addition to change from baseline assessments, change from Screening will also be presented for parameters measured at Screening and later visits.

Continuous data will be summarized using number of subjects (N), mean, standard deviation (SD), median, minimum, and maximum. Summary statistics for continuous data will only include subjects with a non-missing value for the data being summarized.

Geometric mean and geometric coefficient of variation will be produced in addition to the above for pharmacokinetic (PK) data (concentrations and parameters) where appropriate.

Categorical data will be summarized using the number and percent of subjects for each category where appropriate. Missing categories will be used for categorical data when missing data is present. The calculation of percentages will take into account missing responses.

All confidence intervals, statistical tests, and resulting p-values will be reported as 2-sided unless otherwise stated. P-values (when rounded to three decimals) less than or equal to α are reported as "statistically significant".

Means, LS means, and medians will be presented to 1 more decimal place than the recorded data. Where presented, SDs and standard errors (SEs) will be presented to 2 more decimal places than the recorded data. Confidence intervals for parameter estimates will be presented using the same number of decimal places as those recorded for the parameter point estimate.

Alternative methods of analysis of the data may be considered prior to un-blinding should some of the assumptions underlying the proposed analyses not be met. However, the reasons for any departures from the planned approach and methods will be documented fully.

7.2 ANALYSIS SETS

All Subjects: this analysis set will include all subjects who provided informed consent and will include screen fails. This analysis set will be used for reporting of summary data and for data listings.

Enrolled Subjects Set (ESS): this analysis set will include all subjects were enrolled into the PPI run-in period. This analysis set will be used for reporting of summary data.

Full Analysis Set (FAS): this analysis set will include all subjects randomized. Subjects in this set will be analyzed according to the original randomization. The FAS will be the main efficacy analysis set.

Per Protocol Set (PPS): this analysis set is a subset of the FAS. The PPS consists of all subjects who do not violate the terms of the protocol in a way that would impact the study output significantly. All decisions to exclude subjects from the PPS dataset will be made prior to the unblinding of the study. Subjects in this set will be analyzed according to the treatment they actually received. The PPS is an efficacy analysis set and as such exclusion of subjects from the PPS will focus on the effect of violations on the validity of the subject's efficacy data. The PPS analyses will act as a sensitivity analysis to the FAS analyses.

Safety Analysis Set (SAF): this analysis set will include all subjects who have received at least 1 dose of randomized study drug. Subjects in this set will be analyzed according to the treatment they actually received. All safety analyses will be based on the SAF.

The PK set: this set will include all subjects who received randomized treatment and provided at least one PK sample. Subjects in this set will be presented according to the treatment they actually received.

The number of subjects within each analysis set will be summarised by treatment and total.

7.3 DISPOSITION OF SUBJECTS

Disposition will be presented separately for screen failures and for those randomised to the double-blind period. Disposition of all screened subjects (denominator) will be tabulated (count and percent); there will be no inferential analysis of subject disposition data.

Summary of screen failures will be presented for All Subjects. Numbers of subjects failing, and reason for failure, where captured, or continuing into each phase (Screening, PPI run-in, Placebo run-in, double blind randomised phase) and will be presented.

Disposition of all subjects will be tabulated by randomized treatment. The categories will include:

- All randomized subjects (denominator) by site and treatment group.
- Subjects who were randomized but not treated with double-blind study drug, if applicable.

- All subjects who completed double-blind study by site and treatment group.
- All subjects who prematurely discontinued (permanently) study drug by site and treatment group.

Primary reasons for discontinuation of study will be tabulated by treatment and visit

Significant protocol deviations for the FAS will be summarized according to the category of deviation, by site and treatment group.

7.4 Demographic and Baseline Characteristics

Demographic and baseline characteristics will include age (years) at date of informed consent, race, employment status, smoking status, alcohol use, caffeine consumption, gender, height (cm), weight (kg) at Screening, body mass index (BMI) (kg/m2) at Screening, female reproductive status, smoking status, weight gain, night sweats, endoscopy findings and H pylori status at randomisation.

Demographic and baseline characteristics will be summarized by randomised treatment group and total for the ESS. Subjects in the ESS who were screen failures will be included in a 'Run-in Failures' group for presentation purposes. A total randomised column will also be included.

7.5 Medical History and Concurrent Medical Conditions

Summaries of medical history (significant conditions or diseases that stopped at or prior to the time of informed consent) and concurrent medical conditions (significant ongoing conditions or diseases present at the time of informed consent) will be summarized based on the SAF. No inferential statistics will be presented.

Medical history and concurrent medical conditions will be coded using the Medical Dictionary for Drug Regulatory Activities (MedDRA, Version 20 or higher) coding system.

Medical history and concurrent medical conditions will be summarized by system organ class (SOC) and preferred term (PT) for each treatment group. The tables will be sorted in alphabetic order by SOC and in decreasing frequency based on the total number of reports by preferred terms in each SOC. The number and percentage of subjects with any medical history and concurrent medical conditions will be summarized for each SOC and preferred term. The denominator used for calculating the percentages will be the total number of subjects included in each treatment group. For the tables, if a subject reports the same preferred term multiple times, then that preferred term will be counted only once for that subject. Similarly, if a subject reports multiple conditions within the same SOC, then that SOC will be counted only once for that subject in the tables.

7.6 MEDICATION HISTORY AND CONCOMITANT MEDICATIONS

Summaries of medication history and concomitant medications will be based on the SAF. No inferential statistics will be presented.

Medication history includes any medication relevant to eligibility criteria. Concomitant medications include any medication other than study drug (study drug includes esomeprazole administered during the PPI run-in period) taken at any time between the time of informed consent through the end of the study. History of therapy for the eradication of *H. pylori* (eg, triple therapy with PPI + amoxicillin + clarithromycin) will be summarized separately. Any history (Yes/No) and time from completion of such therapy to Screening (within the past 1 year/more than 1 year) will be summarized for SAF by treatment.

Concomitant medications will be classified as follows:

- Medications that started and stopped prior to Visit 4: any medication stopped after time of informed consent and prior to the first dose of double-blind study treatment.
- Medications that started prior to and were ongoing at Visit 4: any medication that started before and was not stopped prior to the first dose of double-blind study treatment.
- Medications that started on or after Visit 4: any medication that started at or after the first dose of double-blind study treatment.
- Medications that were ongoing after Visit 4: any medication that was taken during the double-blind period.

Medication history and concomitant medications will be coded using the WHODrug Version March 2016 or higher. Each of the 4 concomitant medication classifications will be summarized separately.

The tables will present the number and percentage of subjects by therapeutic class and preferred term using the total number of subjects in each treatment group as the denominator. Therapeutic class will be sorted alphabetically and preferred terms are sorted by decreasing frequency based on the total number of subjects.

7.7 STUDY DRUG EXPOSURE AND COMPLIANCE

Study drug exposure and compliance will be summarized by randomised treatment group and total for the SAF. No inferential statistics will be presented.

Exposure and compliance will be summarised for both the run-in and double-blind periods separately. Exposure during the run-in period will be presented for both esomeprazole 40mg and placebo run-in periods separately. Exposure will be summarised as a continuous variable. If any drug accountability data for returned medication is missing it will be assumed that all tablets were taken (ie, amount returned = 0) unless otherwise indicated elsewhere (eg, comment recorded indicating medication lost etc.).

Duration of exposure to drug (days): date of last dose of study drug - date of first dose of study drug + 1. Drug may be esomeprazole run-in, placebo run-in or double blind treatment. If last dose of study drug is missing then the date of the End of Treatment/Early Termination Visit will be used.

Compliance with the study drug regimen will be documented throughout the study and will be calculated from counts of returned tablets and capsules. Study drug compliance for the run-in period will be calculated for the total 6 weeks of run-in (esomeprazole and placebo periods combined). Compliance during the double blind period will also be presented. Compliance will be summarised as a continuous variable and also categorised as follows:

Compliance during the run-in period will be categorised into:

- Compliance <85%.
- Compliance =>85%.

Whilst compliance in the double blind period will be categorised into:

- Compliance <75%.
- Compliance >=75% and <=%133%.
- Compliance>133%.

Compliance will be calculated using:

$$100 \times \left(\frac{\text{Number of capsules taken}}{\text{Expected number of capsules to be taken}} \right)$$

The number of capsules taken will be defined as (total number of capsules dispensed – total number of capsules returned) for each study medication regimen. The expected number of capsules to be taken will be defined as the number of days (last dose day - first dose day + 1) for each study medication regimen of each study period.

Diary compliance will also be summarized. Compliance for the e-Diary is defined as the percentage of scheduled assessments that are completed based on 2 assessments per day (daytime and nighttime), except for the final day of double blind treatment. For randomization, 85% compliance is required and is defined as 12 of 14 assessments completed over the 7-day period (Day -21 to Day-14) prior to the single blind Placebo Run-in Period. Diary compliance will be calculated for the following periods:

- Total run-in period (6 weeks – from Visit 2 to Visit 4).
- Total PPI run-in period (4 weeks – from Visit 2 to Visit 3).
- End of PPI run-in period (1 week – From Day -21 to Day -14).

- Total double blind period (4 weeks – from Visit 4 to Visit 6).

Diary compliance (by treatment and overall) for each period will be summarised as a continuous variable and also categorised into:

- Compliance <85%.
- Compliance =>85%.

7.8 EFFICACY ANALYSIS

The primary and secondary endpoints will be derived from diary data entered by patients into an electronic diary device. The device will present patients with the reflux symptom questionnaire e-Diary (RESQ-eD), an electronic questionnaire developed and validated by Astra Zeneca for use in patients with partial response to PPI. Subjects will be given and trained on the use of the study e-Diary at Visit 2 and re-trained at Visit 3 if necessary once initial eligibility criteria have been confirmed at Visit 3. Subjects will be instructed to complete the e-Diary every morning upon waking, to capture nighttime symptoms, and every evening before going to sleep, to capture daytime symptoms.

Throughout the Run-in Period and Treatment Period, subjects will complete the questionnaire, as well as recording the use of rescue medication in the diary.

The RESQ-eD comprises 13 questions to be answered twice daily. Patients answer each question by selecting one response from a choice of 6 options (did not have, very mild, mild, moderate, moderately severe, and severe).

The 13 questions to be answered regarding daytime symptoms are:

1. Since waking today, how would you rate the intensity of the burning feeling behind your breastbone?
2. Since waking today, how would you rate the intensity of the pain behind your breastbone?
3. Since waking today, how would you rate the intensity of the burning feeling in the center of the upper stomach?
4. Since waking today, how would you rate the intensity of the pain in the center of the upper stomach?
5. Since waking today, how would you rate the intensity of the acid or sour taste in your mouth?
6. Since waking today, how would you rate the intensity of the unpleasant movement of material upwards from the stomach?
7. Since waking today, how would you rate the intensity of your burping (gas coming from the stomach through the mouth)?
8. Since waking today, how would you rate the intensity of your hoarseness?

9. Since waking today, how would you rate the intensity of your cough?
10. Since waking today, how would you rate the intensity of your difficulty swallowing?
11. Since waking today, how would you rate the intensity of the bitter taste in your mouth?
12. Since waking today, how would you rate the intensity of the stomach contents (liquid or food) moving upwards to your throat or mouth?
13. Since waking today, how would you rate the intensity of your heartburn?

Questions 1, 2, 3, 4 and 13 relate to heartburn, whilst questions 5, 6, 11 and 12 relate to regurgitation.

The 13 questions to be answered regarding nighttime symptoms are:

1. During the night time, how would you rate the intensity of the burning feeling behind your breastbone?
2. During the nighttime, how would you rate the intensity of the pain behind your breastbone?
3. During the nighttime, how would you rate the intensity of the burning feeling in the center of the upper stomach?
4. During the nighttime, how would you rate the intensity of the pain in the center of the upper stomach?
5. During the nighttime, how would you rate the intensity of the acid or sour taste in your mouth?
6. During the nighttime, how would you rate the intensity of the unpleasant movement of material upwards from the stomach?
7. During the nighttime, how would you rate the intensity of your burping (gas coming from the stomach through the mouth)?
8. During the nighttime, how would you rate the intensity of your hoarseness?
9. During the nighttime, how would you rate the intensity of your cough?
10. During the nighttime, how would you rate the intensity of your difficulty swallowing?
11. During the nighttime, how would you rate the intensity of the bitter taste in your mouth?
12. During the nighttime, how would you rate the intensity of the stomach contents (liquid or food) moving upwards to your throat or mouth?
13. During the nighttime, how would you rate the intensity of your heartburn?

Questions 1, 2, 3, 4 and 13 relate to heartburn, whilst questions 5, 6, 11 and 12 relate to regurgitation.

A heartburn-free period is defined as a period during which the patient responds 'Did not have' to all questions 1, 2, 3, 4 and 13. A regurgitation free period is defined as a period during which the patient responds 'Did not have' to all questions 5, 6, 11 and 12. For a 24 hour period the patient must be free of symptoms for both the day assessment and the following nighttime assessment (completed the next morning). For analysis purposes, diary entries will be assigned to a study day based on the day the collection interval started for that entry. For example, the Study Day 1 diary entries will include the evening diary completed on Study Day 1 (the collection interval started when the subject awoke on Study Day 1) and the morning entry completed on Study Day 2 (the collection interval started when the subject went to bed on Study Day 1).

For each subject, the percentage of days with neither daytime nor nighttime heartburn during treatment will be calculated using all days with at least 1 morning or evening diary entry during the double blind treatment period (defined as the period starting from Day of first dose of randomised treatment to last day of randomised treatment). For example, if a subject completed at least 1 diary entry on 24 of 27days, but missed both entries on 4 days, 24 days will be used as the denominator in the analysis. On the last day of double blind treatment as the diary will be returned to the site diary entries will not be available for the last day following the last dose of study treatment. This means that if a subject completes 28 days of randomised treatment there should be 27 complete days of diary entries included for consideration for that subject, assuming 100% diary compliance. A similar approach will be taken for daytime and nighttime endpoints and for regurgitation endpoints.

All Efficacy endpoints will be summarised by treatment group and all efficacy data will be listed.

7.8.1 Primary Efficacy Endpoint

Percentage of heartburn-free 24 hour periods (day and night) during 4 weeks of randomized double-blind treatment.

For the assessment of the percentage of heartburn-free days, the Wilcoxon-Mann-Whitney Odds estimator with the 97.5% CI and the Wilcoxon rank-sum test will be used at the 2.5% level of significance for each vonoprazan treatment comparison with esomeprazole to control for multiplicity. This endpoint will be analyzed for both the FAS and PPS. The FAS analysis will be considered the primary analysis and the PPS analysis as a sensitivity analysis.

The percentage of heartburn-free days will be calculated for each subject. When calculating the percentage of heartburn-free days any 24-hour period with no data will be excluded from both the numerator and denominator. A boxplot of the primary endpoint will be presented.

A sensitivity analysis using the FAS will be conducted using the van Elteren test with country as the stratifying factor. An additional analysis will be conducted where a symptom free day will be defined as where both daytime and night time maximum symptom score is very mild for the heartburn questions 1, 2, 3, 4 and 13. This will be analyzed as for the primary endpoint.

Subgroup analyses for the primary efficacy variable will be conducted for the following variables. The subgroup analyses will be conducted on the FAS. Summary statistics will be presented by subgroup level and treatment group for all subgroups. Pairwise comparisons between all treatment groups within each subgroup will be made as for the primary analysis.

- Age (<45, 45-<65, =>65 years old).
- Gender (male, female).
- BMI (<25, 25 to 30, =>30 kg/m2).
- Smoking status (smoker, non-/ex-smoker).
- Alcohol use (drinker, non-/ex-drinker).
- Caffeine use (caffeine user, non-caffeine user).
- H pylori status (positive, negative).
- LA grade at Screening.

7.8.2 Secondary Efficacy Endpoint

Proportion of subjects with ≥1 sustained resolution of heartburn (defined as ≥7 consecutive days without both daytime and nighttime heartburn any time during the 4-week randomized double-blind Treatment Period).

To assess the secondary endpoint, proportion of subjects with ≥1 sustained resolution of heartburn (defined as ≥7 consecutive days without both daytime and nighttime heartburn any time during the 4-week randomized double-blind Treatment Period) a Pearson Chi-square test will be used at the 2.5% level of significance for each Vonoprazan treatment comparison with esomeprazole to control for multiplicity. All subjects with missing data for determination of endpoint status will be considered as a non-responder in the analysis. The difference in proportion and odds ratio for each Vonoprazan treatment comparison with esomeprazole will be estimated and presented with 97.5% confidence intervals (Miettinen-Nurminen confidence intervals). This endpoint will be analyzed for both the FAS and PPS. The FAS analysis will be considered the primary analysis and the PPS analysis as a sensitivity analysis. A sensitivity analysis using the Cochran-mantel-Haenszel test stratifying by country will be conducted using the FAS. A boxplot of the secondary endpoint will be presented.

7.8.3 Additional Efficacy Endpoint(s)

All additional endpoints will be analyzed for the FAS only and tested at the 5% level of significance for each Vonoprazan treatment comparison with esomeprazole. 95% confidence levels will also be presented. Multiplicity will not be adjusted across the remaining endpoints.

If the assumptions underlying planned inferential methods are not adequately met, methods will be amended as needed for appropriate analysis.

Heartburn Endpoints

1. Change from baseline in percentage of heartburn-free 24 hour periods (day and night) during 4 weeks of randomized double-blind treatment, where baseline is defined as the 4 weeks run-in period on esomeprazole.
2. Change from baseline in proportion of subjects with ≥1 sustained resolution of heartburn, where baseline is defined as the 4 weeks run-in period on esomeprazole
3. Percentage of heartburn-free nights during 4 weeks of randomized double-blind treatment.
4. Percentage of heartburn-free days after 4 weeks of randomized double-blind treatment.
5. Proportion of patients with no nighttime heartburn symptoms during the last week of treatment.
6. Proportion of patients with no daytime heartburn symptoms during the last week of treatment.
7. Proportion of patients with no daytime and nighttime heartburn symptoms during the last week of treatment.
8. Percentage of subjects with ≥80% heartburn free days and nights during 4 weeks of randomized double-blind treatment.
9. Percentage of subjects with ≥80% heartburn free nights during 4 weeks of randomized double-blind treatment.
10. Percentage of subjects with ≥80% heartburn free days during 4 weeks of randomized double-blind treatment.
11. Percentage of heartburn free 24 hour periods (day and night) during weeks 1, 2 and week 3 of randomized double-blind treatment.
12. Time to first sustained resolution (defined as 7 consecutive days without daytime and nighttime heartburn).
13. Proportion of subjects who achieve sustained resolution of heartburn (defined as ≥7 consecutive days without daytime and nighttime heartburn) with an onset within 3 days of starting randomized double-blind treatment.
14. Proportion of subjects with resolution of symptoms with an onset within 3 days of starting randomized double-blind treatment and continuing out to the end of week 2 and end of week 4.
15. Maintenance of heartburn-free period – duration of heartburn symptom-free days and nights (24-hour periods) following the first 24-hour period when no symptoms were recorded until the next time symptoms were recorded.

16. Heartburn severity (severity of the most intense episode will be assessed for daytime and nighttime).
17. Percentage of responders, where a responder is defined as having at least 3 more days of not more than mild heartburn symptoms on average per week during the whole double blind treatment period compared to baseline, where baseline is defined as the 1 week run-in period on placebo prior to randomization.

Regurgitation Endpoints

1. Percentage of regurgitation-free nights during 4 weeks of randomized double-blind treatment.
2. Percentage of regurgitation-free days after 4 weeks of randomized double-blind treatment.
3. Percentage of regurgitation-free 24 hour periods during 4 weeks of randomized double-blind treatment.
4. Percentage of subjects with ≥80% regurgitation-free days and nights during 4 weeks of randomized double-blind treatment.
5. Percentage of subjects free from regurgitation ("no regurgitation") during week 1, 2 and 3 of randomized double-blind treatment.
6. Time to first sustained resolution (defined as 7 consecutive days without daytime and nighttime regurgitation).
7. Proportion of subjects who achieve sustained resolution of regurgitation (defined as ≥7 consecutive days without daytime and nighttime heartburn) with an onset within 3 days of starting randomized double-blind treatment.
8. Proportion of subjects with resolution of regurgitation symptoms with an onset within 3 days of starting randomized double-blind treatment and continuing out to the end of week 2 and the end of week 4.
9. Proportion of patients with no daytime and nighttime regurgitation symptoms during the last week of treatment.
10. Maintenance of regurgitation-free period – duration of regurgitation symptom-free days and nights (24-hour periods) following the first 24-hour period when no symptoms were recorded.
11. Regurgitation severity (severity of the most intense episode and will be assessed for daytime and nighttime).

Heartburn and Regurgitation Endpoints

1. Percentage of heartburn and/or regurgitation-free 24 hour periods (day and night) during 4 weeks of randomized double-blind treatment.

Percentage, change from percentage and duration of symptom-free days additional endpoints will be analyzed using the Wilcoxon-Mann-Whitney Odds estimator with the 95% CI and the Wilcoxon rank-sum test will be used at the 5% level of significance for each vonoprazan treatment comparison with esomeprazole. Note that endpoints 8,9,10, 21 and 22 will be analyzed as for proportions. Proportion based additional endpoints will be analyzed using a Pearson Chi-square test at the 5% level of significance for each Vonoprazan treatment comparison with esomeprazole. The difference in proportion and odds ratio for each Vonoprazan treatment comparison with esomeprazole will be estimated and presented with 95% confidence intervals (score confidence intervals).

For time to event data (endpoints 12 and 23) the survivor function will be estimated using the Kaplan-Meier method and the survivor functions between Vonoprazan and Esomeprazole treatment groups will be compared using log-rank tests. A plot of the survivor functions will be presented. The median time to sustained resolution of heartburn from the survivor function will be presented by treatment group, along with the 95% confidence interval.

For severity data the data will be analyzed using the MH chi-square statistic, using standardized midrank scores.

Questions 7, 8, 9 and 10 will not be analyzed.

Rescue medication Endpoint

The following endpoints will be summarized per week by treatment group for rescue mediation:

- Percentage of days without using rescue medication.
- Percentage of nights without using rescue medication.
- Percentage of days and nights without using rescue medication.

The rescue medication percentage endpoints will be analyzed using the Wilcoxon-Mann-Whitney Odds estimator with the 95% CI and the Wilcoxon rank-sum test will be used at the 5% level of significance for each vonoprazan treatment comparison with esomeprazole.

7.9 PHARMACOKINETIC/PHARMACODYNAMIC ANALYSIS

Summary statistics of plasma concentrations of Vonoprazan at the Week 2 and 4 visits and summary PK parameters where available: maximum observed plasma concentration (C_{max}), time to reach C_{max} (t_{max}), area under the plasma concentration-time curve from time 0 to tau (AUC_{τ}). Additional PK parameters may be presented.

Summary statistics of plasma concentrations of Vonoprazan at the Week 2 and 4 visits for CYP2C19 poor or extensive metabolizers (PM/EM) and summary pharmacokinetic parameters where available: maximum observed plasma concentration (C_{max}), time to reach C_{max} (t_{max}), area under the plasma concentration-time curve from time 0 to tau (AUCτ). Additional PK parameters may be presented.

7.10 OTHER OUTCOMES

Nighttime awakenings (eDiary)

The eDiary will include the question "Did you wake up during night (=usual sleeping time)?" in the morning evaluation. The subject response can be 'Yes' or 'No'. The number of nights per week in which the subject woke (response='Yes') will be summarized per week. The % of patients with ≤2 nights of awakenings per week during the last 2 weeks in the double blind treatment period will be reported for each of the 3 treatment arms.

Fatigue (eDiary)

In order to evaluate fatigue, 2 questions will be included in the evening evaluation of the eDiary. The response for both questions is rated on 5-point Likert scale (0 = none, 1 = slight, 2 = medium, 3 = a lot, 4 = severe). The questions are:

1. How much physical fatigue have you experienced today?
2. How much intellectual fatigue have you experienced today?

The number of days per week in which the subject experienced fatigue (response>0) will be summarized per week for each type (physical and intellectual). Additionally the mean level of fatigue per subject per week will be summarized.

Additional Patient Reported Outcome Questionnaires

PROs are recorded at Visit 2 (Day -42, PPI assessment), Visit 3 (Day -14, Off-PPI assessment), Visit 4(Day 1, Randomization) and Visit 6 (Day 28, EOT or early termination). The scores of the PRO questionnaires will be summarized descriptively at each visit including changes from baseline "off-PPI" (visit 4) and changes from baseline PPI response (visit 3) by each treatment group. Data will be assigned to the visit at which it is indicated they were collected in the eCRF. Absolute values and changes from baseline will be summarized per treatment arm. Additionally, the proportion of subjects with clinically significant improvement, when defined for a measure, on HRQoL measures will be summarized.

The following PROs will be recorded:

Pittsburgh Sleep Quality Index (PSQI)

Quality of sleep will be assessed using the PSQI questionnaire. The PSQI questionnaire is a 19-item validated questionnaire completed by patients regarding the previous 1-month period. Items are grouped into seven component scores:

- Subjective sleep quality (1 item).
- Sleep latency (2 items).
- Sleep duration (1 item).
- Habitual sleep efficiency (3 items).

- Sleep disturbances (9 items).
- Use of sleep medication (1 item).
- Daytime dysfunction (2 items).

Each component score is weighted equally on a 0–3 scale, with 3 representing the worse effect, then summed to yield a global PSQI score, which could range from 0 to 21. Component scores as well as global PSQI score are calculated only if all items are answered (eg, a missing question can result in a missing item and a missing global PSQI score). There are five additional questions that are completed by a bed partner if there is one. These are not used in the scoring. Patients with a global score >5 are considered to have poor sleep quality. Only the global score will be summarised. An additional summary table of patients with a global score<=5 and those with a global score >5 at each time point will be presented.

Hospital Depression and Anxiety scale (HADS)

The presence of anxiety levels and depression levels will be assessed using the HADS. The HADS is a widely used scale that was designed for detecting depression and anxiety in an out-patient clinic setting, HADS part A (HADS-A) focuses on anxiety and HADS part D (HADS-D) focuses on depression. HADS-A and HADS-D are 7-item subscales that measure the presence and severity of anxiety and depression symptoms, respectively, on a scale of 0 to 3. Total scores ≤7 indicate no clinically relevant symptoms, 8 to 10 mild symptoms, 11 to 14 moderate symptoms, and ≥15 (maximum 21) indicate more severe symptoms.

The developers recommend that in case of a single missing item in any of the subscales, the score of the subscale to be inferred by using the mean of the remaining six items. If more than one item is missing, then the subscale should be judged as invalid and set to missing. Only the total score (treated as a continuous variable) for each part will be summarised.

Patient Global Impression of Change (PGIC) (not at Visit 2)

The PGIC is a patient-rated instrument that measures change in patient's GERD-related symptoms on a 7-point scale ranging from 1 (very much improved) to 7 (very much worse). Higher scores indicate a change for the worse. For this study, subjects will be asked to rate their symptoms relative to their previous past 4-weeks or 2-weeks. The PGIC will be used to further document the ability of RESQ-eD to detect change as well as clinical meaningfulness of change. The PGIC will be summarised as a continuous variable.

EuroQol-5 dimensions 5 levels (EQ-5D-5L)

The EQ-5D-5L is a validated and reliable generic instrument that was developed as a standardized non-disease-specific instrument describing and evaluating HRQoL. It was intended to complement other forms of HRQoL measures and have the capacity to generate cross-national comparisons for health care evaluations. The questionnaire includes 5 domain items: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Subjects choose the level of

health problems they currently have on each item as “None”, “Slightly”, “Moderate”, “Severe”, or “Unable” and are scored a 1, 2, or 3, respectively. The 5 items allow the calculation of a utility index ranging between -1 (the worst) and 1 (the best). The questionnaire also uses a Visual Analog Scale (VAS). The EQ-5D-5L VAS score is a self-assigned rating of overall health using a 20 cm visual, vertical scale, with a score of 0 as the worst and 100 as best possible health. An increase of ≥0.08 points in the EQ-5D-5L index score represents a meaningful improvement in patient’s HRQoL. The EQ-5D-5L has a recall period of one day. The index score (range -1 -1) and the VAS score (range 0 to 100) will be summarised as continuous variables. Additionally, the proportion of subjects with an increase of ≥0.08 points in the EQ-5D-5L index score at week 4 from each baseline will be summarised.

Patient Assessment of Upper Gastrointestinal Disorders Quality-of-Life Index (PAGI-QOL)

The PAGI-QOl is an upper-GI (UGI)-disorder specific questionnaire and will be used in order to evaluate with more sensitivity the treatment-induced changes in HRQoL. The PAGI-QOL is a validated and reliable instrument used to measure QOL in patients with UGI disorders including GERD, dyspepsia, and gastroparesis. The PAGI-QOL instrument consists of 30 items, each with response options based on a 6-point Likert scale ranging from "None of the time" (0) to "All of the time (5) and with a recall period of the previous 2 weeks. The items are grouped into 5 dimensions:

- Daily Activities (10 items).
- Clothing (2 items).
- Diet and Food Habits (7 items).
- Relationship (3 items).
- Psychological Well-Being and Distress (8 items).

A score per dimension as well as a total score can be calculated. PAGI-QOL subscale scores are calculated by taking the mean of the items in each subscale after reversing item scores. PAGI-QOL subscale scores range from 0 (lowest QoL) to 5 (highest QoL). In case of missing data, the half-scale rule is applied (ie, the subscale score is calculated when 50% or less of the items are missing for that scale; if more than 50% of the items within a subscale are missing, the score is set to missing). The PAGI-QOL total score is calculated by taking the mean of the corresponding subscales. If a subscale score is set to missing the PAGI-QOL total score is then set to missing. Only the total score will be summarised.

Patient Assessment of Upper Gastrointestinal Disorders Symptom Severity Index (PAGI-SYM)

Changes in UGI symptoms will be measured with the PAGI-SYM, a validated and reliable instrument used to measure symptom severity in patients with UGI disorders including GERD,

dyspepsia, and gastro paresis. The PAGI-SYM measures UGI symptom severity over the 2 weeks preceding the visit. The PAGI-SYM contains 20 items, each on a 6-point Likert scale ranging from None (0) to Very severe (5). The PAGI-SYM includes 6 subscales:

- Heart burn/regurgitation (7 items).
- Post-prandial fullness/early satiety (4 items).
- Nausea/vomiting (3 items).
- Bloating (2 items).
- Upper abdominal pain (2 items).
- Lower abdominal pain (2 items).

The PAGI-SYM subscale scores are calculated by taking the mean of the items in each subscale; the subscale scores vary from 0 (none or absent) to 5 (very severe). The half-scale rule is applied for missing data (ie, the subscale score is calculated using the mean of non-missing items; when more than 50% of items are missing, the score is set to missing). The total score is calculated by taking the mean of the subscales. If a subscale score is missing, the PAGI-SYM total score is set to missing. Only the total score will be summarised.

PAGI-SYM will be used to identify patients that are more likely to be functional patients. PAGI-SYM will also be used to identify meaningful change in RESQ-eD.

Exploratory Analyses

CCI

7.11 SAFETY ANALYSIS

Safety analyses include AEs, clinical laboratory values, vital signs, and other safety parameters. All safety summaries will be based on the SAF. The analysis of safety data will be restricted to descriptive statistics only unless otherwise specified. AE tables will include individual treatment columns and a column for both Vonoprazan arms combined (Total Vonoprazan) as well as Total over all 3 arms.

7.11.1 Adverse Events

See the protocol for definitions of pretreatment events (PTE), adverse events (AEs) and serious AEs (SAEs). Treatment-emergent adverse events (TEAEs) will be defined as any AE that occurs after the first dose of study drug during the randomized double blind period (for the purposes of assigning an AE to this period the start of the randomized double blind period will be from time of randomization on Day 1) and up to until 1 week after completion or withdrawal from the study (safety follow-up call). PTEs and AEs occurring during the run-in phase will be listed and summarized by SOC and PT.

There are no AEs of special interest defined for the study.

The MedDRA dictionary, version 20 or higher, will be used to code all AEs reported during the trial by SOC and PT.

All TEAE summaries will be arranged in alphabetical order of SOC then by descending frequency (subject frequency) of the preferred term (PT) within the Total Vonoprazan treatment group.

AEs with missing intensity will be listed as such in the AE listings, however, will be noted with an asterisk and summarized as severe in summary tables. If the relationship of an event is missing, the relationship for the event will be considered to have been related. In the cases where a subject has multiple AEs with the same SOC or PT the AE with the maximum intensity or strongest relationship will be summarized. When calculating the frequency and percentage of subjects who reported TEAEs, a subject will be counted only once for each SOC or PT when multiple TEAEs are coded to the same SOC or preferred term. Thus, if a subject has two distinct AEs, each of which corresponds to a distinct preferred term but both of which correspond to the same SOC, then that subject will be counted once at that SOC subject-count summary level and once at each of the two preferred-term subject-count summary levels.

The number and percentage of subjects with TEAEs will be summarized in several different tables:

- Overview of TEAEs (including number and percentage of patients and events).
- All TEAEs by SOC and PT.
- All TEAEs by SOC.
- All TEAEs by PT.
- Most frequent TEAEs by SOC and PT (occurring in ≥5% of subjects within one treatment group).
- Most frequent non-serious TEAEs by SOC and PT (occurring in ≥5% of subjects within one treatment group).
- Relationship of TEAEs to Study Drug by SOC and PT (not related, related).

- Drug-Related TEAEs by SOC and PT.
- Intensity of TEAEs by SOC and PT.
- Intensity of Drug-Related TEAEs by SOC and PT.
- TEAEs Leading to Study Drug Discontinuation by SOC and PT.
- Serious TEAEs by SOC and PT.
- Pretreatment Events by SOC and PT.
- Pretreatment Serious Events by SOC and PT.
- Run-in Adverse Events by SOC and PT.
- Serious Run-in Adverse Events by SOC and PT.
- Listing of Deaths.

Subject mappings will also be produced for:

- TEAEs.
- TEAEs leading to study drug discontinuation.
- Serious TEAEs.

A list of PTEs by subject number and MedDRA coding will be presented separately. PTEs will be summarized by SOC and PT

7.11.2 Clinical Laboratory Evaluations

Clinical laboratory tests and urinalysis will be performed at Screening, Visit 4 (Randomization, Day 1), Visit 5 (Day 14) and Visit 6 (EOT or early termination). Laboratory values will be summarized, for absolute and change from baseline (Visit 4) values and will be presented by parameter and visit for the SAF. Values will be mapped to a scheduled visit by the use of visit windows, if a subject has multiple values within a particular visit window, the most extreme result will be used for summary. Additionally a final visit summary will be provided, summarizing the last recorded values. All laboratory test parameters will be displayed in individual subject data listings in standard international (SI) units. Shifts in laboratory test values will be presented as cross-tabulations (baseline versus each post-baseline visit) of numbers of subjects with low, normal and high values relative to the normal range used at the central laboratory. This classification will be based on the low, normal and high alert flags reported by the central laboratory. Shift tables will be produced for all clinical laboratory tests with reference ranges. Individual results for clinical laboratory tests that meet the predefined criteria for markedly abnormal laboratory values (MAV) will be summarized in tables. MAV criteria are given in Appendix.

Visit Windows for Safety Laboratory Parameters			
visit	**Baseline**	**Week 2**	**Week 4/ET**
Day	<=1	14	28
window	<=1	2-21	22 - 42

Liver function parameters ALT, AST and total bilirubin will be reported separately. In addition to the standard shift table, the number of subjects with AST, ALT and total bilirubin values > upper limit of normal (based on the central laboratory defined normal range), >2 ULN and >3ULN will also be summarized by visit and overall (double blind period). Additionally the number of subjects with AST or ALT >3 ULN and with total bilirubin >2ULN will also be summarised by visit and overall (double blind period).

CCI [redacted]

The following tables will be produced:

- Summary of Laboratory Test Results and Change from Baseline by Study Visit.
- Number and Percent of Subjects in Categories of Urine Laboratory Parameters by Visit.
- Summary of Shifts of Laboratory Test Results.
- Number and Percent of Subjects With MAVs of Laboratory Parameters.
- Number and Percent of Subjects With Elevated Liver Enzyme Laboratory Parameters.

Subject mappings will also be produced for:

- The Number and Percent of Subjects With MAVs of Laboratory parameters.
- Subjects With Elevated Liver Enzyme Laboratory Parameters.

To obtain information on H. pylori infection status at baseline a 13C or 14C urea breath test will be performed at Randomization by the local laboratory. A stool test or blood test to measure the infection of H. pylori may be performed at those centers which do not have the facilities to perform a H pylori breath test procedure. The results of the H. Pylori tests will be listed separately from the other laboratory parameters. A summary of H. Pylori status (number and percent of Subjects by H Pylori status) will be included in the baseline and demographic tables.

7.11.3 Vital Signs

Vital signs, including body temperature (°C), sitting blood pressure (mmHg) and pulse (bpm), and weight (kg) will be measured at Visit 1 (Screening), Visit 4 (Day 1, Randomization), and

Visit 6 (day 28, End of Treatment) or Early Termination. Height (cm) will also be measured at Screening. Measures taken at Visit 4 (Day 1) will be considered as baseline values. Vital signs will be summarized by visit and treatment group for the SAF. Change from baseline (Visit 4, Day 1) will also be summarized for Visit 6 (Day 28). Data will be assigned to the visit at which it is indicated they were collected in the eCRF. No statistical analysis will be performed on vital signs data. Individual results for vital signs parameters that meet the predefined criteria for MAVs will be summarized in tables. MAV criteria are given in Appendix.

The following tables and subject mappings will be produced:

- Summary of Vital Signs Parameters and Change from Baseline by Visit.
- Number and Percent of Subjects With MAVs of Vital Signs Parameters.
- Subject Mappings for the Number and Percentage of Subjects With MAVs of Vital Signs Parameters.

Listings of all vital signs will be provided for all subjects. MAVs will also be flagged in the listings.

7.11.4 12-Lead ECGs

12-lead ECG (comprising interpretation (within normal limits, abnormal but not clinically significant, or abnormal and clinically significant), heart rate, RR interval, PR interval, QT interval and QRS interval) will be measured at Visit 1 (Screening), Visit 4 (Day 1, Randomization), and Visit 6 (day 28, End of Treatment) or Early Termination. Measures taken at Visit 4(Day 1) will be considered as baseline values. ECG will be summarized by visit and treatment group for the SAF. Change from baseline (Visit 4, Day 1) will also be summarized for Visit 6 (Day 28). Data will be assigned to the visit at which it is indicated they were collected in the eCRF. The corrected QT using the Fredericia correction ($QTcF = QT/\sqrt[3]{RR}$) will also be calculated and presented. No statistical analysis will be performed on vital signs data. Individual results for ECG parameters that meet the predefined criteria for MAVs will be summarized in tables. MAV criteria are given in Appendix.

The following tables and subject mappings will be produced:

- Summary of ECG Parameters and Change from Baseline by Visit.
- Number and Percent of Subjects With MAVs of 12-lead ECG Parameters.
- Number and Percent of Subjects in Various Categories of ECG Parameters.
- Subject Mappings for the Number and Percentage of Subjects With MAVs of ECG Parameters.
- Subject Mappings for the Subjects in Various Categories of ECG Parameters.

Listings of all vital signs will be provided for all subjects. MAVs will also be flagged in the listings.

7.11.5 Other Observations Related to Safety

Weight (kg) will be measured at Visit 1 (Screening), Visit 4 (Day 1, Randomization), and Visit 6 (day 28, End of Treatment) or Early Termination. Measures taken at Visit 4 (Day 1) will be considered as baseline values. Weight will be summarized by visit and treatment group for the SAF. Change from baseline (Visit 4, Day 1) will also be summarized for Visit 6 (Day 28). Data will be assigned to the visit at which it is indicated they were collected in the eCRF. No statistical analysis will be performed.

The following tables will be produced:

- Summary of Weight and Change from Baseline by Visit.

A listing of weight will be provided for all subjects, including height(cm) and BMI (kg/m2) at Screening.

A physical examination will be conducted at Visit 1 (Screening), Visit 4 (Day 1, Randomization), and Visit 6 (day 28, End of Treatment) or Early Termination. The physical examination at Visit 4 (occurring prior to first dose of double blind treatment) will be considered as baseline. This examination will consist of the following body systems: eyes; ears, nose, throat; cardiovascular system; respiratory system; gastrointestinal system; dermatologic system; extremities; musculoskeletal system; nervous system; lymph nodes; and other. All subsequent physical examinations should assess clinically significant changes from the assessment from the baseline examination. Physical examination data will only be listed; any clinically significant changes are captured and reported as AEs.

7.12 INTERIM ANALYSIS

Not applicable

7.13 CHANGES IN THE STATISTICAL ANALYSIS PLAN

No changes

8.0 REFERENCES

Study Protocol, dated 25 November 2015

Study Protocol Amendment 1, dated 15 February 2016 (non-substantial)

Study Protocol Amendment 2, dated 26 March 2016 (non-substantial)

9.0 APPENDIX

9.1 CLINICAL LABORATORY TESTS

Hematology	Serum Chemistry	Urinalysis
Hematocrit Hemoglobin Platelet count RBC count Reticulocyte count WBC count (with differential) Erythrocyte indices (MCV, MCH, MCHC) **Coagulation Tests** Prothrombin time Partial thromboplastin time International normalized ratio	ALT Albumin Alkaline phosphatase Amylase AST Blood urea nitrogen Calcium Chloride Cholesterol CK C-reactive protein (CRP) (a) Gastrin tests (b) γ-Glutamyl transferase Glucose Lactate dehydrogenase Magnesium Pepsinogen I/II(c) Potassium Phosphate Serum creatinine Sodium Standard iron panel (serum iron, TIBC, serum ferritin) (d) Total bilirubin Direct bilirubin (e) Total protein Triglycerides Uric acid	Bilirubin Glucose Hemoglobin Ketones Leucocytes Nitrites pH Protein Specific gravity Urobilinogen **Microscopy** for abnormal urinalysis result (RBCs, WBCs, casts, crystals, organisms, and epithelial cells)

Other:

Serum	Urine	Breath
Hepatitis panel, including HBsAg and anti-HCV Thyroid-stimulating hormones (TSH)(f) <u>Female subjects only</u> Beta hCG (for pregnancy, female subjects of childbearing potential) Follicle-stimulating hormone (FSH) if menopause is suspected (g)	<u>Female subjects only</u> hCCGI (for pregnancy, female subjects of childbearing potential) <u>All subjects</u> Drug screen including: amphetamines (including methamphetamine), barbiturates, benzodiazepines, cannabinoids, cocaine, opiates, methadone and phencyclidine (h)	Helicobacter pylori (i)

MCH=mean corpuscular hemoglobin, MCHC=mean corpuscular hemoglobin concentration, MCV= mean corpuscular volume, RBC=red blood cell, TIBC= total iron binding capacity, WBC=white blood cell.
(a) CRP will be collected at Visit 4 and Visit 6(or Early termination).
(b) Gastrin and Gastrin-17 will be performed.
(c) This includes pepsinogen I, pepsinogen II and pepsinogen I/II ratio.
(d) Standard iron panel will be collected at Visit 1 and Visit 6(or Early termination).
(e) Assess direct bilirubin only if total bilirubin >1.5×ULN.
(f) TSH will be checked at Visit 4 and Visit 6(or Early termination).
(g) Follicle-stimulating hormone will be conducted at investigator's discretion to determine the postmenopausal status of women whose duration of (consecutive) amenorrhea is borderline or open to doubt and where the investigator believes the subject to be menopausal by history.
(h) The central laboratory will confirm any positive drug screen results.
(i) *Helicobacter pylori* breath test will be performed at Randomization via a urea breath test (C13 or C14 urea test) at a local laboratory. A stool test or a blood test may be performed at those centers which do not have the facilities to perform a H pylori breath test procedure.

9.2 CRITERIA FOR MARKEDLY ABNORMAL VALUES

Hematology

Parameter	MAV Criteria	
	Lower Criteria	Upper Criteria
Red Blood Cells ($\times 10^{12}$/L)	<0.8×LLN	>1.2×ULN
White Blood Cells ($\times 10^{9}$/L)	<0.5×LLN	>1.5×ULN
Hemoglobin (g/L)	<0.8×LLN	>1.2×ULN
Hematocrit (%)	<0.8×LLN	>1.2×ULN
Platelets (×109/L)	<75	>600
Neutrophils (%)	<0.5×LLN	>1.5×ULN
Eosinophils (%)	-	>2×ULN
Basophils (%)	-	>3×ULN
Monocytes (%)	-	>2×ULN
Lymphocytes (%)	<0.5×LLN	>1.5×ULN

Serum Chemistry

Parameter	MAV Criteria	
	Lower Criteria	Upper Criteria
ALT (U/L)	-	>3×ULN
ALP (U/L)	-	>3×ULN
AST (U/L)	-	>3×ULN
GGT (U/L)	-	>3×ULN
Total Bilirubin (µmol/L)	-	>34.2
Direct Bilirubin (µmol/L)	-	>2×ULN
CK (CPK) (U/L)	-	>5×ULN
Albumin (g/L)	<25	-
Total Protein (g/L)	<0.8×LLN	>1.2×ULN
Creatinine (µmol/L)	-	>177
BUN (mmol/L)	-	>10.7
Uric Acid (mmol/L)	-	>0.773
Total Cholesterol (mmol/L)	-	>7.72
Triglycerides (mmol/L)	-	>2.5×ULN
Glucose (mmol/L)	<2.8	>19.4
Potassium (mmol/L)	<3.0	>6.0
Sodium (mmol/L)	<130	>150
Magnesium (mmol/L)	<0.5	>1.2
Calcium (mmol/L)	<1.75	>2.88
Inorganic Phosphorus (mmol/L)	<0.52	>2.00
Chloride (mmol/L)	<75	>126

Vital Signs

Parameter	MAV Criteria	
	Lower Criteria	Upper Criteria
Body Temperature (C)	<35.6	>37.7
Systolic Blood Pressure (mmHg)	<85	>180
Diastolic Blood Pressure (mmHg)	<50	>110
Pulse (bpm)	<50	>120

12-lead ECG

Parameter	MAV Criteria	
	Lower Criteria	Upper Criteria
Heart Rate (bpm)	<50	>120
QT Interval (msec)	<=50	>=460
QTcF Interval (msec)	<=50	>=500 OR (>=30 change from baseline and >=450)

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'UTC')
PPD	Biostatistics Approval	22-Jan-2019 03:59 UTC
	Clinical Science Approval	25-Jan-2019 08:01 UTC