Title: A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Phase 2b Study to Evaluate the Efficacy and Safety of Twice-Daily Oral Administration of a Peripherally Acting Dopamine Receptor D2/D3 Antagonist, TAK-906 for the Treatment of Adult Subjects with Symptomatic Idiopathic or Diabetic Gastroparesis

NCT Number: NCT03544229

SAP Approve Date: 02 August 2021

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.

# STATISTICAL ANALYSIS PLAN

## STUDY NUMBER: TAK-906-2002

A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Phase 2b Study to Evaluate the Efficacy and Safety of Twice-Daily Oral Administration of a Peripherally Acting Dopamine Receptor D2/D3 Antagonist, TAK-906 for the Treatment of Adult Subjects with Symptomatic Idiopathic or Diabetic Gastroparesis

**PHASE 2b**

Version: Final

Date: 02 August 2021

**Prepared by:**

PPD

Based on:

Protocol Version: Amendment 08

Protocol Date: 13 July 2020

## 1.1 Approval Signatures

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

## 2.0 TABLE OF CONTENTS

## LIST OF IN-TEXT TABLES



## LIST OF IN-TEXT FIGURES



## LIST OF APPENDICES

## 3.0 LIST OF ABBREVIATIONS

| | |
|---|---|
| AE | adverse event |
| ALT | alanine aminotransferase |
| ANMS GCSI-DD | American Neurogastroenterology and Motility Society Gastroparesis Cardinal Symptom Index-Daily Diary |
| ANCOVA | analysis of covariance |
| ANOVA | analysis of variance |
| AST | aspartate aminotransferase |
| BID | twice daily |
| BLQ | below the limit of quantification |
| BMI | body mass index |
| BUN | blood urea nitrogen |
| CCK | Cholecystokinin |
| CGI-I | Clinical Global Impression Scale–Global Improvement Scale |
| CGI-S | Clinical Global Impression Scale–Severity of Illness Scale |
| CI | confidence interval |
| $C_{trough}$ | observed concentration at the end of a dosing interval |
| CPAP | Clinical Pharmacology Analysis Plan |
| CRP | C-reactive protein |
| CSSR | clinician symptom severity rating |
| DA | dopamine |
| DG | diabetic gastroparesis |
| ECG | electrocardiogram |
| eCRF | electronic case report form |
| eGFR | estimated glomerular filtration rate |
| ePRO | electronic patient reported outcomes |
| EPS | extrapyramidal symptoms |
| FAS | full analysis set |
| GE | gastric emptying |
| GEBT | gastric emptying breath test |
| GGT | γ-glutamyl transferase |
| GI | Gastrointestinal |
| GIP | glucose-dependent insulinotropic polypeptide |
| GP | Gastroparesis |
| HbA1c | glycosylated hemoglobin |
| IA | Interim analysis |
| ICF | Informed Consent Form |
| ICH | International Conference on Harmonisation |
| IDMC | Independent Data Monitoring Committee |
| IG | idiopathic gastroparesis |
| IVRS | Interactive Voice Response System |
| LDH | lactate dehydrogenase |
| LLN | lower limit of normal |

| | |
|---|---|
| LOCF | last observation carried forward |
| LS | least squares |
| MAR | missing at random |
| MAV | markedly abnormal value |
| MCS | Mental Component Summary |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MI | Multiple Imputation |
| MMRM | mixed model for repeated measures |
| MNAR | missing not at random |
| OC | observed case |
| OTE | Overall Treatment Effect |
| PAGI-QoL | Patient Assessment of Upper Gastrointestinal Disorders-Quality of Life Index |
| PAGI-SYM | Patient Assessment of Upper Gastrointestinal Disorders-Symptom Severity Index |
| PCS | Physical Component Summary |
| PD | pharmacodynamics |
| PGI-I | Patient Global Impression of Improvement |
| PGI-S | Patient Global Impression of Severity |
| PGx | pharmacogenomics |
| PK | pharmacokinetics |
| PO | oral administration or orally |
| PRO | patient-reported outcome |
| PT | preferred term |
| PTE | pretreatment events |
| PYY | Peptide YY |
| QOL | quality-of-life |
| QTcF | QT interval with Fridericia correction method |
| RBC | red blood cell |
| SAE | serious adverse event |
| SAP | statistical analysis plan |
| SD | standard deviation |
| SDB | standard database |
| SE | Standard error |
| SF12 | Short Form-12 Health Survey |
| SOC | system organ class |
| TEAE | treatment-emergent adverse event |
| TLGs | tables, listings, and graphs |
| ULN | upper limit of normal |
| WBC | white blood cell |
| WHODrug | World Health Organization Drug Dictionary |

# 4.0 OBJECTIVES

## 4.1 Primary Objectives

To assess the efficacy of treatment with various dose levels of TAK-906 in adult subjects with gastroparesis compared with placebo during the 12 weeks of treatment.

## 4.2 Secondary Objectives

To evaluate the safety and tolerability of TAK-906 doses compared with placebo during the 12 weeks of treatment.

## 4.3 Additional Objectives

- To evaluate the PK of TAK-906 in subjects with gastroparesis.
- To evaluate the effect of TAK-906 on quality of life compared with placebo.
- To evaluate the potential relationship between changes in C-reactive protein (CRP), glucose-dependent insulinotropic peptide (GIP), and peptide tyrosine tyrosine (PYY) and clinical and motility responses to TAK-906.
- To evaluate the effect of TAK-906 on Gastric Emptying Breath Test (GEBT).

## 4.4 Study Design

This is a global multicenter, randomized, double-blind, placebo-controlled, parallel-group, phase 2b study to evaluate the efficacy and safety of BID PO capsule administration of a peripherally acting DA receptor $D_2/D_3$ antagonist, TAK-906, in adult male and female subjects aged 18 to 85 years, inclusive, with symptomatic IG or DG. Gastroparesis is a disorder of the stomach characterized by delayed gastric emptying, in the absence of mechanical obstruction.

The study is designed to characterize the dose-response relationship to support dose selection for evaluation of subjects to be enrolled in the planned phase 3 studies.

There was a substantial protocol amendment after the COVID-19 pandemic (Protocol Amendment 8, 13 July 2020), where
1) the 5 mg arm dropped
2) modifications were made in data collection to alleviate site burden and based on agency recommendations.

Subjects will attend the study site for a screening/consent visit (Visit 1, Day -35), when their consent will be obtained, and general eligibility to participate in the study will be reviewed. After signing the informed consent form, subjects will have a blood sample taken to assess their laboratory values. Subjects will discontinue all excluded medications and return to the site 3 to 17 days later to perform a 4-hour C-Spirulina GEBT (Visit 2, if required).

In Protocol Amendment 8, subjects who have confirmed of delayed gastric emptying confirmed by an accepted diagnostic test method (ie, scintigraphy, GEBT, or wireless motility capsule) that

is documented in the subject's medical records prior to screening, are not required to undergo the Visit 2 GEBT, and may proceed directly to Visit 3.

Subjects who are not taking any medications at the screening visit that require washout may attend the clinic for Visit 2 earlier than 2 weeks and only after laboratory results have been reviewed by the investigator. An 8-hour fast is required before the GEBT. Subjects will then return within 7 days after the GEBT visit to confirm their eligibility based on the gastric emptying criteria (Visit 3, Day -14).

Once eligibility is confirmed by these criteria at this visit, subjects will be instructed on the use of an electronic patient reported outcomes tool (ePRO) for collecting the daily symptoms of gastroparesis, daily dosing and meal time recordings, and capture of rescue medication use. The instrument used to capture symptom data will be the ANMS GCSI-DD along with a bloating severity scale. Subjects will record their symptoms once daily (in the evening) and will return to the clinic approximately 2 weeks later to assess their gastroparesis symptom eligibility based on the ANMS GCSI-DD (Visit 4, randomization). Subjects need to be compliant with completing the Daily Symptom Diary, defined as ≥80% diary completions, during the 2-week symptom assessment period.

Medication should be taken 1 dose in the morning approximately 1 hour before the first meal of the day and another dose in the evening approximately 1 hour before the main last meal of the day, at a regular dose interval.

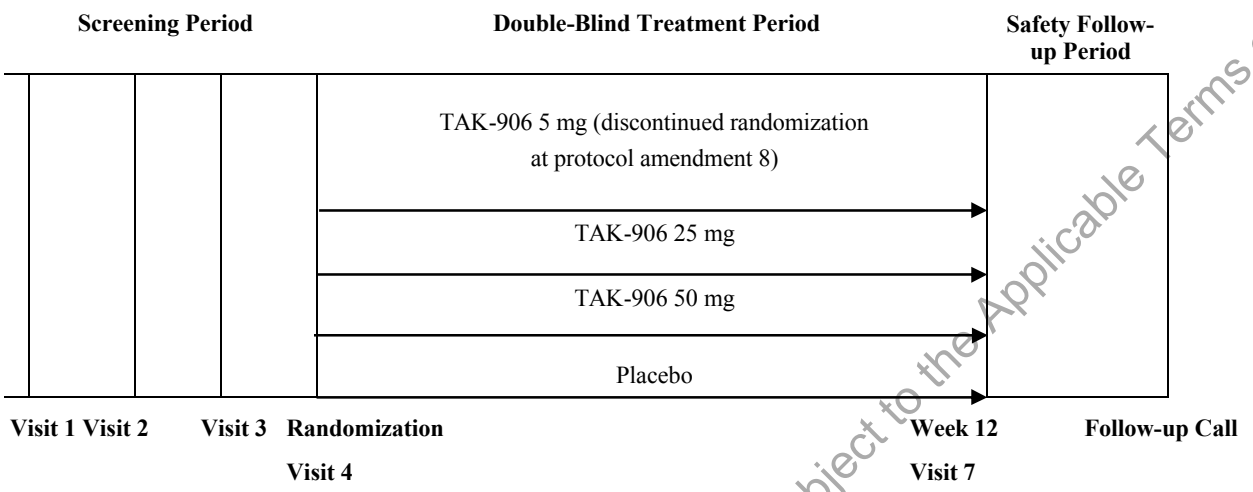
Prior to Protocol Amendment 8, Subjects was randomized 1:1:1:1 into 1 of 4 treatment groups: TAK-906 maleate (TAK906M abbreviated as TAK-906) 5, 25, and 50 mg capsule BID or matching placebo BID, stratified by IG and DG (minimum 30 subjects per indication per arm).

In Protocol Amendment 8, the randomization into 5 mg arm for TAK-906 was discontinued while the randomization into the rest of the treatment groups remains unchanged, ie, subjects will be randomized 1:1:1 into TAK-906 25 mg and 50 mg capsule BID or placebo BID.

Subjects who withdraw prematurely will be seen by the physician for an early termination visit as soon as possible and will continue completing the Daily Symptom Diary until this termination visit using the ePRO device. They will then be contacted for a follow-up safety call to check if any adverse events occurred during this follow-up period.

A schematic of the study design is included as Figure 4.a.

**Figure 4.a Schematic of Study Design**




## 5.0 ANALYSIS ENDPOINTS

### 5.1 Primary Endpoint

- Change from baseline to Week 12 of the Treatment Period in the American Neurogastroenterology and Motility Society Gastroparesis Cardinal Symptom Index-Daily Diary (ANMS GCSI-DD) composite score (nausea, early satiety, upper abdominal pain, and postprandial fullness).

### 5.2 Secondary Endpoints

- Proportion of subjects with at least 50% reduction from Baseline in ANMS GCSI-DD composite score at Week 12.
- Change from baseline to Week 12 of the Treatment Period in the ANMS GCSI-DD nausea symptom score.
- Change from baseline to Week 12 of the Treatment Period in the ANMS GCSI-DD early satiety symptom score.
- Change from baseline to Week 12 of the Treatment Period in the ANMS GCSI-DD postprandial fullness symptom score.
- Change from baseline to Week 12 of the Treatment Period in the ANMS GCSI-DD upper abdominal pain symptom score.
- Change from baseline to Week 12 of the Treatment Period in the ANMS GCSI-DD recorded vomiting frequency.

- Change from baseline to Week 12 of the Treatment Period in the ANMS GCSI-DD overall severity of gastroparesis symptoms score.
- Change from baseline to Week 12 of the Treatment Period in the bloating severity scale score.
- Change from baseline to Week 12 of the Treatment Period in the ANMS GCSI-DD total score (nausea, early satiety, upper abdominal pain, postprandial fullness, bloating, and vomiting).
- Proportion of symptomatic weeks (weeks with average composite symptom score assessed as >mild [ANMS GCSI-DD Score ≥2]) during 12 weeks of treatment.
- Change from baseline to Week 12 of the Treatment Period in the Patient Assessment of Upper Gastrointestinal Disorders-Symptom Severity Index (PAGI-SYM) total score.

### 5.3 Safety and Tolerability Endpoints

Safety and tolerability will be evaluated using the following general assessments:

- Treatment-emergent AEs.
- Vital signs.
- 12-lead ECGs.
- Clinical laboratory parameters (hematology, clinical chemistry, and urinalysis).

### 5.4 Additional Endpoints

- Percentage of symptomatic days (days with symptoms assessed as >mild [ANMS GCSI-DD Score ≥2]) using the ANMS GCSI-DD during 12 weeks of treatment.
- Proportion of patients with ≥20% reduction in ANMS GCSI-DD composite score from Baseline in the last 6 consecutive weeks of the study.
- Change from baseline to Week 12 of the treatment period in the ANMS GCSI-DD core symptom score (nausea, early satiety, upper abdominal pain, postprandial fullness, and vomiting).
- Change from baseline to Week 12 of the Treatment Period in Clinician Symptom Severity Rating form.
- Change from baseline to Week 12 of the Treatment Period in Patient Assessment of Upper Gastrointestinal Disorders Quality-of-Life Index (PAGI-QOL) total score.
- Differences at Week 12 in Overall Treatment Effect (OTE) Scales.
- Change from baseline to Week 12 in Clinical Global Impression Scale–Severity of Illness Scale (CGI-S).

- Differences at Week 12 in Clinical Global Impression Scale–Global Improvement Scale (CGI-I).
- Change from baseline to Week 12 in Patient Global Impression of Severity (PGI-S) both version 1 (Collected daily) and version 2 (collected over the past 7 days).
- Differences at Week 12 in Patient Global Impression of Improvement (PGI-I) and Patient Global Impression of Change (PGI-C)
- TAK-906 predose (trough) plasma concentration at Week 12.
- Change from baseline to Week 12 of the Treatment Period in Short Form-12 (SF12) Health Survey subscale scores.
- Use of rescue medication during the 12-week Treatment Period.
- Assess change in CRP, GIP, and PYY from baseline to Week 12.
- Change from baseline to Week 4 and 12 in GEBT time to gastric half emptying ($t_{1/2}$).

## 6.0 DETERMINATION OF SAMPLE SIZE

In order to manage the likelihood of statistical, clinical, and operational success of this study in the setting of the COVID-19 pandemic, the sample-size considerations are revisited in Protocol Amendment 8. The statistical testing is revised from 2-sided (with a significance level of 5%) to 1-sided (with a significance level of 5%), the planned sample size is reduced by 10 subjects per arm and randomization into the 5 mg dose arm is to be discontinued.

The sample-size rationale for the resultant reduced sample size and updated 1-sided statistical testing is as follows: Assuming a SD of 1.25 and a true difference of 0.625 between TAK-906 active dose and placebo in the change from baseline in the ANMS GCSI-DD composite score, and a 15% dropout rate, a total of 205 subjects is sufficient to achieve approximately 80% power to detect the treatment effect of 0.625 at a 1-sided 5% significance level. The total of 205 subjects consists of approximately 25 subjects who have already been randomized to receive 5 mg capsule BID prior to terminating randomization of subjects into the 5 mg dose arm, and 60 subjects per treatment groups of placebo, TAK-906 25 mg, and TAK-906 50 mg.

## 7.0 METHODS OF ANALYSIS AND PRESENTATION

### 7.1 General Principles

Categorical data will be summarized as the number and percentage of subjects in each category. Percentages will be reported to 1 decimal place.

Continuous data will be summarized using descriptive statistics, including the number of subjects, mean, standard deviation (SD), median, minimum, and maximum. The coefficient of variation (%CV) and geometric mean will be included in the summary of continuous data where indicated. Arithmetic means, geometric means, and medians will be presented to 1 more decimal

place than the recorded data, and SDs will be presented to 2 more decimal places than the recorded data, where appropriate.

All statistical tests will be 1-tailed at $\alpha$=0.05 level for statistical significance unless otherwise stated. Where applicable, 90% confidence intervals (CIs) corresponding to a parameter estimate will be presented using the same number of decimal places as the parameter estimate and p-values will be rounded to 3 decimal places prior to assessment of statistical significance.

The logistic regression models based on maximum likelihood approach will be used to analyze binary efficacy endpoint at each visit separately as default. In case quasi-complete separation issue exists at a specific visit where the maximum likelihood estimate from the logistic regression model becomes questionable, the Firth's penalized likelihood approach (ie, model /FIRTH option for Proc Logistic in SAS) will be used instead for that visit.

Study-related raw data for enrolled subjects, including derived data, will be presented in data listings where indicated. The actual day relative to the first dose will be presented, where applicable.

All statistical analyses will be performed using the SAS System Version 9.4 or a later version. SF-12 Mental Component Summary (MCS) and Physical Component Summary (PCS) will be calculated based on PRO Core® (version 2, 17-Dec-2020) a software from Quality Metric (Maruish, 2012).

In case of irregular data entry, the algorithms below will be used to identity the ANMS GCSI-DD score for analysis. All ANMS GCSI-DD scores will be listed and flagged if the score is excluded from analysis.

1. If a subject has data entry for ANMS GCSI-DD scores entered one day earlier before the planned data entry date or at least two days after the planned data entry date, the ANMS GCSI-DD scores will not be used for analysis. Only the ANMS GCSI-DD scores entered on the same day as planned data entry date or one day after (within one calendar day) will be used for analysis. Data points entered after first dose will not be used for deriving baseline.
2. If a subject has two entries for the ANMS GCSI-DD scores for the same day, the data entry that will be used for analysis is determined as below:
    - If both entries are before the evening*, the entry closer to the evening* will be used for analysis.
    - If one entry is before and one entry is after the evening*, the entry after the evening* will be used for analysis because it encompasses longer time interval for the respective day.
    - If both entries are after the evening*, the latter entry will be used for analysis as it encompasses longer time interval for the respective day (and as long as the entry is within one calendar day).

*as per the study protocol, patients were advised to complete the dairy in the evening at 5.00 p.m

### 7.1.1 Definition of Study Day, Baseline and Study Visit Windows

A windowing convention will be used to determine the value for analysis variables for a given study visit that applies to observed data.

Study Day 1 is defined as the date of the first dose of study drug, as recorded on the electronic case report form (eCRF). Other study days are defined relative to Study Day 1. Study days prior to the first dose of study drug will be calculated as: [date of interest – date of first dose of study drug]. Study days on or after the first dose of study drug will be calculated as: [date of interest – date of first dose of study drug + 1].

For each visit, a window will be defined such that the lower and upper bounds of each window is generally the midpoint between 2 consecutive study visits. The visit windows and applicable study day ranges are presented below in Table 7.a.

More than 1 result for a study endpoint measurement may be obtained in a visit window. In such an event, the result with the date closest to the scheduled visit day will be used. In the event of 2 observations equidistant to the scheduled visit day, the later of the observations will be used.

In the safety data summary, the treatment period 'End of Treatment Visit' values will be defined irrespective of falling in a particular window. Hence, the windowed Week 12 value may be different than the End of Treatment Visit value.

**Table 7.a Visit Windows**

| Analysis /Nominal Visit Week | Nominal Visit Day | ANMS GCSI-DD | CSSR, SF12 | OTE (b), CGI-I, PGI-I | GEBT (c) | PAGI-SYM, PAGI-QoL, CGI-S, PGI-S, ECG, Vital Sign, Lab |
|---|---|---|---|---|---|---|
| Baseline | -1 (a) | ≤0 | ≤1 | NA | ≤1 | ≤1 |
| 1 | 7 | 1—7 | | | | |
| 2 | 14 | 8—14 | | | | |
| 3 | 21 | 15—21 | | | | |
| 4 | 28 | 22—28 | | | 2—56 | 2—42 |
| 5 | 35 | 29—35 | | | | |
| 6 | 42 | 36—42 | | | | |
| 7 | 49 | 43—49 | | | | |
| 8 | 56 | 50—56 | | | | 43—70 |
| 9 | 63 | 57—63 | | | | |
| 10 | 70 | 64—70 | | | | |
| 11 | 77 | 71—77 | | | | |
| 12 | 84 | 78—84 | ≥ 2 | ≥ 2 | ≥ 57 | ≥ 71 |

(a) Nominal Visit Day -1 corresponds to the date(s) of the Baseline visit for each assessment. For some assessments, baseline is collected on Study Day 1 before study drug administration. Daily Collected ANMS GCSI-DD Scores will be derived into Analysis Week Values.

(b) Both OTE subject and OTE physician scales.

(c) Collection of GEBT for post baseline assessment was removed in Protocol Amendment 8.

In general, the baseline value for a variable is defined as the last non-missing observation prior to the first dose of double-blind study medication (visit date/time ≤first dose date/time), including the screening value, if necessary. The baseline value for ANMS GCSI-DD scores is defined in Section 7.8.2.

Adverse events that start more than 30 days after the last dose of double-blind study medication (start date – last dose date >30) will be listed, but excluded from the summaries and analyses. For efficacy and other safety measurements, data that are obtained more than 7 days after the last dose of double-blind study medication (visit date – last dose date >7) will be listed, but excluded from summaries and analyses.

The data listings will display Analysis Visits or Nominal Visits where applicable.

### 7.1.2 Grouping of Centers

No adjustment of study centers in the statistical analyses will be made.

## 7.2 Analysis Sets

- **Safety Analysis Set**: The safety analysis set will include all subjects who were randomized and received at least 1 dose of double-blind study medication. In safety summaries, subjects will be analyzed according to the treatment they received.

  If a subject receives more than 1 treatment, the actual treatment is defined as the one that is used most frequently. If two or more treatments are tied in frequency of use, the actual treatment will be the one with highest dose level.

- **Randomized Set**: The randomized set will include all subjects who were randomized. In summaries based on randomized set, subjects will be analyzed by the treatment to which they were randomized.

- **Full Analysis Set (FAS)**: The FAS will include all subjects who were randomized, received at least 1 dose of study drug, and have a baseline value for assessment of primary endpoint. In FAS efficacy summaries, subjects will be analyzed by the treatment to which they were randomized.

- **Per-Protocol Set (PPS)**: The PPS will include all FAS subjects who had no major protocol violations that could significantly impact the clinical outcomes. If more than 5% of the total subjects in the FAS have major protocol violations, analyses based on the PPS will be performed for the primary efficacy variable only. Subjects with major protocol violations will be identified as part of the blinded data review prior to the unblinding of subject's treatment assignment, and will be listed by study center and subject number. The categories of major protocol violations include:

  a) informed consent was not obtained

  b) Not meeting inclusion/exclusion criteria

     - inclusion criteria of 3, 4, 5, 6, 9, 13
     - exclusion criteria of 6, 12, 13, 14, 18, 23, 30
     - exclusion criteria 34 for Protocol Amendment 8 enrolled subjects, exclusion criteria 36 for subjects enrolled prior to Protocol Amendment 8

  c) Receiving incorrect study medication

  d) Low study drug compliance (<80%)

  e) Study medication exposure less than 8 weeks

     - ((last dose date – first dose date +1)/7) <8)

  f) Treatment blind broken

  g) Use of rescue medications is used 3 days BID per week for more than 2 consecutive weeks. If end date of rescue medication is missing, the end date of rescue medication will be imputed by last date of rescue medication from the ePRO device.

## 7.3 Disposition of Subjects

Study Information, including date of first subject signing Informed Consent Form (ICF), date of first/last study drug, date of last subject's last visit/contact, date of last subject's last procedure for collection of data for primary endpoint, Medical Dictionary for Drug Regulatory Activities (MedDRA) Version, World Health Organization Drug Dictionary (WHODrug) Version, and SAS Version, will be tabulated.

Disposition of all screened subjects will be summarized descriptively, including a summary of the number of screened subjects, the number of subjects eligible/not eligible for randomization, and the primary reason for ineligibility for randomization. The number of screen failures and their characteristics will also be summarized. The inclusion and exclusion data will be listed by subject.

Disposition for all randomized subjects will be summarized by treatment group and overall. Disposition categories include:

- Number of randomized subjects by country and site.
- Number of subjects randomized but not dosed, subjects completing or prematurely discontinuing study drug along with the primary reason for study drug discontinuation, and subjects completing or not completing all study visits along with the primary reason for discontinuation of study visits.

Important/Significant protocol deviations will be summarized based on the Randomized Set. Reasons for exclusion from the PPS will be summarized based on the FAS. All protocol deviations will be listed, separately. COVID-19 related Protocol Deviations will be summarized separately.

The impact due to COVID-19 will be summarized using the number and percentage of subjects in each category of impact by treatment group and overall based on all randomized subjects:

- Subjects with at least one alternative contact method used due to COVID-19.
- Subject with alternative contact method used due to COVID-19 by study visit.
- Subjects with assessments done via alternative contact method by study visit.

Meanwhile, a listing of all participants who were affected by the COVID-19 related study disruption will be provided by including unique subject number identifier, investigational site, and a description of how the individual's participation was altered.

The analysis sets defined in Section 7.2 will be summarized.

## 7.4 Demographic and Other Baseline Characteristics

Demographic and baseline characteristics including sex, age, age category (18-64 years, 65-84 years, 85 years and over), race, height, weight, BMI (<25, 25-<30, >=30), and smoking status will be listed and summarized for each treatment group and overall. The demographic and baseline characteristics summary will be based on all randomized subjects.

Disease history characteristics including disease population (DG, IG), prior diagnosis of GP (Yes, No), time since symptom onset (month), confirmation of absence of upper GI obstruction though imaging (Yes, No), prior diagnostic test(s) for GP (Scintigraphy, GEBT, Wireless Motility Capsule, Other), and test results (Delayed Gastric Emptying, Normal Gastric Emptying) will be listed and summarized for each treatment group and overall. The disease history summary will be based on randomized set.

Baseline values for efficacy parameters (ANMS GCSI-DD, PAGI-SYM, and Clinician Symptom Severity Rating, PGI-S, CGI-S, PAGI-QoL, MCS SF-12, PCS SF-12, and GEBT $t_{1/2}$) will also be presented for each treatment group, overall TAK-906 treated group based on the FAS.

All individual demographic and baseline data will be listed by treatment, and subject number.

## 7.5 Medical History and Concurrent Medical Conditions

Medical history refers to significant conditions/diseases that resolved at or before Screening (time of informed consent). Concurrent medical conditions are those significant ongoing conditions/diseases present at Screening (time of informed consent).

Medical history and concurrent medical conditions will be coded using Medical Dictionary for Regulatory Activities (MedDRA) latest version and will be summarized by treatment group and overall using System Organ Class (SOC) and MedDRA preferred term. The table will include number and percentages of subjects, and will be sorted in alphabetical order by SOC. Within an SOC, PTs are sorted in decreasing frequency based on the total number of subjects. A subject will only be counted once within a particular class even if he/she has multiple conditions/symptoms. Summaries will be based on all randomized subjects.

All medical history and concurrent medical condition data will be listed by treatment, study center and subject number.

## 7.6 Medication History and Concomitant Medications

The medication history and concomitant medications are defined as follows:

- Medication history refers to the medication that the study subjects stopped taking at or within 90 days prior to time of first dose.
- Concomitant medication is defined as medication that the study subjects continued taking or took from time of first dose through end of study:
  - Concomitant medication that started prior to and was ongoing at first dose date (ie, start date < first dose date, and stop date ≥ first dose date).
  - Concomitant medication that started after first dose (ie, start date ≥ first dose date).
  - Concomitant medication taken during the study (ie, start date and stop dates are between first and last dose date).
  - Concomitant medication that started post last dose (ie, start date >last dose date).

If start date and stop date are missing, medication will be assumed to occur both prior and concomitantly.

Medication history and concomitant medications will be coded using the latest version of the World Health Organization (WHO) Drug Dictionary and summarized by giving the number and percentage of subjects by preferred term within each therapeutic class, with therapeutic class sorted in alphabetical order and preferred term sorted in decreasing frequency based on the total number of subjects. The number of subjects with medications in each selected therapeutic class will also be presented. If a subject report taking 2 drugs belonging to the same class, he/she will only be counted once within that class. Summaries of medication history and concomitant medication will be based on all randomized subjects.

In addition, number and percentage of subjects who took antinausea/antivomiting rescue medication will be summarized.

All prior and concomitant medications will be listed by treatment, study center and subject number. Rescue medications will also be listed by treatment, study center, and subject number. Rescue medications from ePRO device and EDC will be listed, separately.

### 7.7 Study Drug Exposure and Compliance

The summary of study drug exposure and compliance will be based on the safety analysis set.

Duration of exposure to double-blind study medication is defined as (date of last dose – date of first dose +1). Treatment duration will be summarized by duration category in weeks (<1 week, ≥1 to <4 weeks, ≥4 to <8 weeks, ≥8 to <12 weeks, ≥12 weeks) and the number of subjects in each duration category by treatment group. Treatment duration (weeks) will also be summarized as a continuous variable using descriptive statistics.

For subjects who returned all dispensed bottles, the following equation will be used to calculate the percent of study drug compliance:

(number of capsules dispensed – number of capsules returned)/[(date of last dose – date of first dose +1)×2]×100%.

For subjects who did not return all bottles that were dispensed, the dosing data from ePRO device will be used to impute the number of capsules taken for calculating the percent of study drug compliance during the period of time for which the bottle(s) was not returned:

[sum of (number of capsules dispensed – number of capsules returned, at the corresponding visit from IRT) for the visits where dispensed bottles were returned + sum of number of capsules taken from ePRO device for the visits where dispensed bottles were not returned]/ [(date of last dose – date of first dose +1)×2]×100%.

If the missing dispensed bottle is not the last bottle, count the number of capsules taken from the dispense date to the day before next dispense date in ePRO device.

If the missing dispensed bottle is the last bottle, count the number of capsules taken from the last dispense date to date of last capsule taken in ePRO device.

For each treatment group, study medication compliance will be summarized by compliance category (<80%, 80%-100%, >100%) and the number of subjects in each compliance category.

Study medication compliance will also be summarized as a continuous variable using descriptive statistics for each treatment group.

All study drug administration and accountability data will be listed by treatment, study center, and subject number. The following variables will be listed: subject identifier, first and last dose dates, medication identification number, date dispensed and returned, number of capsules dispensed and returned, and percent compliance. Study drug data collected from ePRO will be listed.

## 7.8 Efficacy Analysis

The analyses and summaries for efficacy will be based on the FAS.

The primary, secondary, and additional efficacy endpoints for this study are presented in Table 7.b. All derived endpoints (indicated by 'D') in this table will be accompanied by a listing of intermediate data for supporting the derivation including but not limited to weekly scores and subscale calculations, severity scores, and missing data imputations.

The composite score, total score, and core symptom score will be graphed.

**Table 7.b Primary, Secondary and Additional Efficacy Endpoints**

| Endpoint (1) | Parameter | Description | Variable Type (2) |
|---|---|---|---|
| Primary | ANMS GCSI-DD_COMP | Composite score (nausea, early satiety, postprandial fullness, upper abdominal pain). | c/ D |
| Secondary | ANMS GCSI-DD_TOT | Total score (nausea, early satiety, postprandial fullness, upper abdominal pain, bloating, and vomiting). | c/ D |
| | ANMS GCSI-DD_CS | Core symptom score (nausea, early satiety, upper abdominal pain, postprandial fullness, and vomiting). | c/ D |
| Secondary | ANMS GCSI-DD_NA | Nausea symptom score | c |
| Secondary | ANMS GCSI-DD_ES | Early satiety symptom score | c |
| Secondary | ANMS GCSI-DD_PF | Postprandial fullness symptom score | c |
| Secondary | ANMS GCSI-DD_UAP | Upper abdominal pain symptom score | c |
| Secondary | ANMS GCSI-DD_VM | Vomiting symptom frequency score | c |
| Secondary | ANMS GCSI-DD_GL | Overall severity of GP symptom | c |
| Secondary | ANMS GCSI-DD_RESP1 | ANMS GCSI-DD response defined as ≥50% decrease from baseline in ANMS GCSI-DD composite score at week 12 | b/ D |
| Secondary | ANMS GCSI-DD_WK | Proportion of symptomatic weeks (weeks with average composite symptom score assessed as worse than mild [weekly ANMS GCSI-DD_COMP≥2]) during 12 weeks of treatment. | c/ D |
| | ANMS GCSI-DD_RESP2 | ANMS GCSI-DD response defined as ≥20% | b/ D |

## Table 7.b Primary, Secondary and Additional Efficacy Endpoints

| Endpoint (1) | Parameter | Description | Variable Type (2) |
|---|---|---|---|
| | | decrease from baseline in ANMS GCSI-DD composite score in the last 6 consecutive weeks of the study. | |
| | ANMS GCSI-DD_DAY | Percentage of symptomatic days (days with symptoms assessed as worse than mild [daily ANMS GCSI-DD_COMP≥2]) during 12 weeks of treatment. | c/ D |
| Secondary | BLOATING | Bloating severity scale score | c |
| Secondary | PAGI-SYM_TOT | Mean of PAGI-SYM subscale scores | c |
| | PAGI-SYM_NAVM | Mean of Nausea/vomiting subscale score (nausea, retching, vomiting) | c/ D |
| | PAGI-SYM_PFES | Mean of Postprandial fullness/early satiety subscale score (stomach fullness, not able to finish a normal-sized meal, feeling excessively full after meals, loss of appetite) | c/ D |
| | PAGI-SYM_BL | Mean of Bloating subscale score (bloating, stomach or belly visibly larger) | c/ D |
| | PAGI-SYM_UAP | Mean of Upper abdominal pain subscale score (upper abdominal pain, upper abdominal discomfort) | c/ D |
| | PAGI-SYM_LAP | Mean of lower abdominal pain subscale score (lower abdominal pain, lower abdominal discomfort) | c/ D |
| | PAGI-SYM_HBRE | Mean of Heartburn/regurgitation subscale score (heartburn during the day, heartburn when lying down, feeling of discomfort inside your chest during the day, feeling of discomfort inside your chest at night, regurgitation or reflux during the day, regurgitation or reflux when lying down, bitter, acid or sour taste in your mouth) | c/ D |
| | CSSR_TOT | Clinician symptom scale rating total score | c/ D |
| | PAGI-QoL_TOT | Mean of PAGI-QoL subscale scores | c/ D |
| | SF-12 | SF-12 subscale scores (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health) | c/ D |
| | MCS SF-12 | Mental Component Summary SF-12 | c/P |
| | PCS SF-12 | Physical Component Summary SF-12 | c/P |
| | GEBT $t_{1/2}$ | GEBT time to gastric half emptying | c/ D |

**Table 7.b Primary, Secondary and Additional Efficacy Endpoints**

| Endpoint (1) | Parameter | Description | Variable Type (2) |
|---|---|---|---|
| | OTE_S | Overall treatment effect (OTE) scale rated by subject | o |
| | OTE_C | Overall treatment effect (OTE) scale rated by clinician | o |
| | CGI-I | Clinician Global Impression of improvement | o |
| | CGI-S | Clinician Global Impression of severity | o |
| | PGI-S (3) | PGI severity | o |
| | PGI-C (4) | PGI change | o |
| | PGI-I (4) | PGI improvement | o |

(1) Primary and Secondary Endpoints are identified. The rest are additional Endpoints.
(2) b = binary; c = continuous, o = ordinal / D =derived, P=PRO Core software used
(3) PGI-S has two versions (PGI-S Version 1 prior to Protocol Amendment 08: recall period is the same day when the form was filled out; PGI-S Version 2 under Protocol Amendment 08: recall period over the past 7 days).
(4) PGI-I was replaced with PGI-C (PGI-Improvement: collected daily prior to Protocol Amendment 08; PGI-Change: collected under Protocol Amendment 08).
(5) CGI-I, CGI-S, CSSR_TOT and OTE_C will be summarized when responded by qualified personnel on or after they were trained. Otherwise, data will be available in the listings only.

### 7.8.1 Endpoints Derivation based on ANMS GCSI-DD Daily Scores

The ANMS GCSI-DD **individual symptom** scores are as follows:

- The scores for nausea, early satiety, postprandial fullness, and upper abdominal pain range from 0 to 4.
- The open-ended vomiting frequency assesses the number of vomiting episodes during the day.
- The patient global rating of the overall severity of GP symptom ranges from 0 to 4.

The ANMS GCSI-DD **daily composite** and **total** scores are calculated as follows:

- The daily **composite score** is calculated by summing the scores on the 4 symptom items (nausea, early satiety, postprandial fullness, and upper abdominal pain) and then dividing by 4, that is the number of items within the composite score. Thus, the maximum daily composite score could be (4 symptoms × maximum score 4 divided by 4) = 16/4 = 4. The ANMS GCSI-DD daily composite score can range from 0 to 4 with higher scores reflecting greater symptom severity.

- The daily **total score** is calculated by summing the scores on each of the 5 symptom items in ANMS GCSI-DD (nausea, early satiety, postprandial fullness, upper abdominal pain, and vomiting) plus the bloating severity item and then dividing by 6. When calculating the total score, the vomiting frequency will be scored from 0 to 4 (where 0 = no vomiting and 4 = four or more episodes of vomiting), similar to the scoring of the other individual symptom items. The maximum total symptom score could be (6 symptoms × maximum score 4 divided by 6) = 24/6 = 4. The ANMS GCSI-DD daily total score can range from 0 to 4 with higher scores reflecting greater symptom severity.
- The daily **core symptom score** is calculated by summing the scores on each of the 5 symptom items in ANMS GCSI-DD (nausea, early satiety, postprandial fullness, upper abdominal pain, and vomiting) and then dividing by 5. When calculating the core symptom score, the vomiting frequency will be scored from 0 to 4 (where 0 = no vomiting and 4 = four or more episodes of vomiting), similar to the scoring of the other individual symptom items. The maximum core symptom score could be (5 symptoms × maximum score 4 divided by 5) = 20/5 = 4. The ANMS GCSI-DD daily core symptom score can range from 0 to 4 with higher scores reflecting greater symptom severity.

The ANMS GCSI-DD **weekly scores** for individual items, composite, total, and core symptom will be derived from the daily scores calculated based on electronic diaries completed by the patients. These scores for baseline and each week will be calculated as follows:

- **Baseline value**: the baseline value is the average of daily scores from the most recent 7 consecutive days (Day -7 to Day -1) before the first dose date of the baseline visit.
- The score for each postbaseline week is the average of the 7 daily scores within the targeted week relative to the first dose date. For example, Week 1 score will be the average of the 7 daily scores from Day 1 to Day 7; Week 2 score will be the average of the 7 daily scores from Day 8 to Day 14; and so on for duration of the study.
- General rules for dealing with missing individual items or daily scores are described in Section 7.8.5.

### 7.8.2 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline to Week 12 of the Treatment Period in the weekly ANMS GCSI-DD composite score.

#### *7.8.2.1 Analysis of Primary Efficacy Endpoint*

The change from baseline in weekly ANMS GCSI-DD composite score will be analyzed using a mixed-effects model for repeated measures (MMRM) with treatment, week, subject disease population (DG/IG), treatment-by-week interaction as fixed effects, baseline and baseline score-by-week interaction as covariates, with an unstructured variance covariance matrix. The estimates of the least squares mean (LSM) difference between TAK-906 dose groups and placebo at each week and their 90% CIs will be reported. In particular, the LSM differences between TAK-906 dose groups and placebo at Week 12 will be used to evaluate the treatment

effects. The restricted maximum likelihood (ie, REML) method will be used to estimate the covariance parameters and Kenward-Roger approximation will be used to estimate the denominator degrees of freedom for the tests of fixed effects resulting from the model.

If the model does not converge, other variance covariance matrix structures will be considered in the following order:

1. Fit MMRM model with unstructured covariance using Singular statement in Proc Mixed (eg, singular=**1e-11).**
2. Kiernan, Tao, and Gibbs from SAS institute (2012) pointed out that non-convergence may happen when the default convergence criteria in SAS are too tight (eg, parameter estimates have a small magnitude and the default relative convergence criteria cannot be satisfied) and setting one's own criteria can help to obtain convergence.
3. Fit MMRM model with unstructured covariance using PARMS statement in Proc Mixed (initial values will be obtained based on Fisher scoring algorithm).
4. Kiernan, Tao, and Gibbs (2012) pointed out that the default starting values in SAS might not work for a particular dataset and using the PARMS statement to specify a different set of starting values can help to obtain convergence. The Fisher scoring algorithm (via the SCORING option of the PROC MIXED statement) will be used to obtain the initial values of covariance parameters.
5. Use the no-diagonal factor analytic structure (via the TYPE=FA0(*T*) option of the REPEATED statement, where *T* is the total number of time points).
6. Lu and Mehrotra (2010) pointed out that no-diagonal factor analytic structure effectively performs the Cholesky decomposition of the covariance matrix and is numerically more stable.
7. Use the following alternative covariance structures sequentially (from generous to parsimonious):

Toeplitz, Variance Components, Autoregressive (1), and Compound Symmetry.

The same variance covariance matrix structure will be used for the corresponding sensitivity analysis.

The following sensitivity analyses will be conducted for the primary efficacy endpoint:
*To deal with subjects who received rescue therapy during the treatment period:*

1. A sensitivity analysis will be done by considering ANMS GCSI-DD scores reported on and after start of rescue medication through 48 hours after the end date of rescue medication as missing at random (MAR). The rescue medication is assumed not to have any effect on the ANMS GCSI-DD scores 48 hours after the end date of rescue medication.
2. A sensitivity analysis will be done by imputing ANMS GCSI-DD scores reported on and after start of rescue medication through 48 hours after the end date of rescue medication by considering the Nausea symptom level as the worst score (ie, 4). Because the protocol

specified use of antinausea/antivomiting rescue medication is permitted for moderate nausea and/or vomiting CTCAE Grade 2 (for nausea, oral intake decreased without significant weight loss, dehydration or malnutrition; for vomiting; outpatient IV hydration required) the rescue medication is assumed to primarily affect Nausea or Vomiting symptoms and not to have any effect on the these symptoms 48 hours after the end date of rescue medication.

The timing of 48 hours for the two sensitivity analysis above was determined by 5 times the half-life of rescue medications allowed in the study.

*To deal with subjects who have missing values:*

1. As a sensitivity analysis to the MAR assumption in the primary analysis described above, a pattern mixture model (PMM) with the control-based imputation will be applied to handle missing values. The strategy for implementing control-based pattern imputation using regression-based multiple imputation (MI) for monotone missingness (Ratitch 2011 and Li 2019) can be summarized in 4 steps of the imputation process as follows. Details are in Appendix E Details for Pattern Mixture.

If more than 5% of total subjects in the FAS are excluded, the primary analysis will also be performed based on the PPS.

Subgroup Analyses

For the primary efficacy endpoint, subgroup analyses by:

- Subject disease population (DG, IG) (the MMRM for subgroup analysis will not include disease population as a fixed effect),
- Age (≤65, >65),
- Sex (female, male),
- Baseline composite score (≤median, >median),
- Geographical Region (North America, Europe, Asia),

will be performed. The treatment groups will be compared within each subgroup.

### 7.8.3 Secondary Efficacy Endpoints

Change from baseline in weekly ANMS GCSI-DD total score, ANMS GCSI-DD individual symptom scores (nausea, early satiety, upper abdominal pain, and postprandial fullness), recorded vomiting frequency, overall severity of gastroparesis symptoms score, and bloating severity scale score, will be analyzed using similar method (MMRM) for primary efficacy endpoint.

The recorded vomiting frequency greater or equal to 60 is deemed as clinically implausible, and will be set to missing for analysis. The derivation for total ANMS GCSI-DD score will be based on this vomiting frequency (setting clinically implausible number as missing) capped at 4 (User Manuel for the ANMS GCSI-DD).

The estimates of the least squares mean (LSM) difference between TAK-906 dose groups and placebo at each week and their 90% CIs will be reported. In particular, the LSM differences between TAK-906 dose groups and placebo at Week 12 will be used to evaluate the treatment effects in each endpoint, respectively.

The proportion of subjects with at least 50% reduction from Baseline in the weekly ANMS GCSI-DD composite score will be estimated by logistic regression adjusting for baseline score, subject disease population, and treatment by week. The missing values will be handled in 3 ways by last observation carried forward (LOCF), observed case (OC) and non-responder imputation (NRI).

- For LOCF imputation, the last non-missing post-baseline values would be carried forward to impute the weekly ANMS GCSI-DD composite score.
- For the OC analysis, only includes subjects who have non-missing ANMS GCSI-DD composite score, there will not be any imputation.
- For NRI imputation, subject with missing ANMS GCSI-DD composite score would be handled as non-responder.

Percentage of symptomatic weeks (weeks with average composite symptom score assessed as >mild [ANMS GCSI-DD weekly composite score ≥2]) during 12 weeks of treatment will be analyzed using ANOVA with treatment, subject disease population as fixed factors. If the weekly ANMS GCSI-DD composite score is missing, that week is considered a symptomatic week.

Change from baseline in PAGI-SYM total and subscale scores will be analyzed using similar method as the primary efficacy endpoint. The estimates of the least squares mean (LSM) difference between TAK-906 dose groups and placebo at each week and their 90% CIs will be reported. In particular, the LSM differences between TAK-906 dose groups and placebo at Week 12 will be used to evaluate the treatment effects in each endpoint, respectively.

The PAGI-SYM contains 20 items grouped into 6 subscale scores. A 6-point Likert response scale, ranging from 0 (none) to 5 (very severe), is used for rating the severity of each symptom.

Normality assumption will be assessed for the analysis of weekly ANMS GCSI-DD composite scores, total score, and core symptom scores. If a significant departure from the normality assumption is detected, a generalized estimating equations (GEE) approach will be applied.

The GEE analysis model will be similar to the MMRM for the analysis of primary efficacy endpoint (with treatment, week, subject disease population (DG/IG), treatment-by-week interaction as fixed effects, baseline and baseline score-by-week interaction as covariates, with an unstructured working covariance matrix as default). If GEE model with unstructured working covariance matrix fails to converge, use alternative working covariance structures sequentially: Autoregressive (1), then Compound Symmetry.

### 7.8.4 Additional Efficacy Endpoints

The proportion of subjects with ≥20% reduction in ANMS GCSI-DD composite score from baseline in the last 6 consecutive weeks of the study will be estimated by logistic regression model adjusting for baseline score, subject disease population, and treatment. The missing value will be handled in 3 ways by (1) last observation carried forward (LOCF), (2) observed case (OC) and (3) non-responder imputation (NRI).

For the observed case analysis, only includes subjects who have the length of treatment period >= 6 weeks.

- If a subject consistently met ≥20% reduction in ANMS GCSI-DD weekly composite score from baseline for the last 6 consecutive weeks, the subject will be a responder.
- If a subject does not consistently meet the ≥20% reduction in ANMS GCSI-DD weekly composite score from baseline for the last 6 consecutive weeks, the subject will be a non-responder.

For the LOCF imputation, all subjects would be included.

- The last non-missing post-baseline ANMS GCSI-DD weekly composite score prior to the missing weeks will be carried forward to impute the missing values.
- Based on the imputed data, if a subject met ≥20% reduction in ANMS GCSI-DD weekly composite score from baseline for last 6 consecutive weeks, the subject will be a responder. Otherwise, the subject will be a non-responder.

For the non-responder imputation, all subjects would be included. The responder and non-responder will be derived as follows:

- The subject will be considered as non-responder if, subject
  - Participated in study less than 6 weeks or
  - Had missing values for ANMS GCSI-DD composite weekly score in the last 6 consecutive weeks of the study or
  - Didn't meet the ≥20% reduction in ANMS GCSI-DD weekly score for the last 6 consecutive weeks of the study.
- If a subject consistently met ≥20% reduction in ANMS GCSI-DD weekly composite score from baseline for last 6 consecutive weeks, the subject would be a responder.

Percentage of symptomatic days (days with symptoms assessed as >mild [ANMS GCSI-DD composite score ≥2]) using the ANMS GCSI-DD during 12 weeks of treatment will be analyzed using ANOVA with treatment, subject disease population as fixed factors. The day with missing daily composite score will be considered symptomatic day.

Change from baseline in CSSR total score (CSSR total score will be an average of all the items, in which each item ranges from 0 (none) to 5 (very severe)) and SF-12 subscale (based on norm-

based scores) scores will be analyzed using ANCOVA with treatment and subject disease population as fixed factors, and baseline score as covariate.

As a supplementary analysis, the CSSR based on nominal visit will be analyzed using ANCOVA with treatment and subject disease population as fixed factors, and baseline score as covariate.

Change from baseline in SF-12 PCS and MCS will be analyzed using ANCOVA with treatment and subject disease population as fixed factors, and baseline score as covariate. SF-12 derived sub-scales, and sub-sales that are norm-based scores will all be listed.

Change from baseline in weekly ANMS GCSI-DD core symptom score, weekly PAGI-QoL total, and weekly subscale scores, and weekly GEBT $t_{1/2}$ will be analyzed using similar method as the primary efficacy endpoint. The estimates of the least squares mean (LSM) difference between TAK-906 dose groups and placebo at each week and their 90% CIs will be reported. In particular, the LSM differences between TAK-906 dose groups and placebo at Week 12 will be used to evaluate the treatment effects in each endpoint, respectively.

OTE-C and OTE-S will be summarized for the TAK-906 dose groups and placebo by the landing question and its subsequent question.

CGI-I, CGI-S, PGI-I, and PGI-C scores will be tabulated (count and percentage) by visit for TAK-906 dose groups and placebo.

As a supplemental analysis, the CGI-S scores based on the nominal visit will be tabulated (count and percentage) by visit for TAK-906 dose groups and placebo.

PGI-S will be summarized based on which versions of the PGI-S subjects answered: PGI-S (version 1), PGI-S (version 2), and PGI-S (mixed versions 1 and 2). The PGI-S tables will have total, and subgroups by Disease Type at Randomization. All data will be listed.

For the clinician-based rating scores (CGI-I, CGI-S, CSSR_TOT and OTE_C), the analysis will only include rating scores that were evaluated by qualified and trained PI/SI/Staff per MedAvante training requirements as shown in table below. All datapoints will be listed.

| Qualified/ Trained? | Evaluator Category | Evaluated by trained and qualified PI/SI/Staff |
|---|---|---|
| PI/SI trained (or IM trained) and qualified | 1 | Y |
| Staff trained and qualified | 2 | Y |
| PI/SI - not qualified or NOT trained | 3 | N |
| Staff qualified but not trained | 4 | N |
| Staff not qualified and not trained | 5 | N |

*Raters that aren't principle investigator or sub-investigator would be considered staff.

### 7.8.5 Missing Data

For some reasons, subjects may not complete the full task as requested, or subjects used irregular procedure and/or questionable effort. In such cases, the data may be considered missing, unreliable and/or invalid.

The general rule when individual items are missing from a multiple-item assessment is as follows:

- Composite, core symptom score, or total ANMS GCSI-DD score will be calculated as (sum of non-missing item scores) / (number of non-missing items).
- If more than 20% of the items are missing, the composite/total score/core symptoms score will be set to missing. For example, if there is one item missing for a daily average, then the denominator will be the number of non-missing items (eg, for total score, if one of the items are missing, then divide by 5 rather than 6).
- If 3 or less days of diary data are available for the baseline or a postbaseline week, that weekly score will be considered missing (ie, each week would need at least 4 of the 7 days).

Missing date imputation for medication and adverse events is detailed in Appendix F Convention for Missing dates for Adverse Events/Concomitant Medication/Procedures. These imputation rules will be derived to create flags, but in the listings display actual collected.

## 7.9 Pharmacokinetic/Pharmacodynamic Analysis

### 7.9.1 Pharmacokinetic Analysis

No formal noncompartmental PK analyses will be performed on concentration-time data. Individual concentration-time data will be included in summary table and listings. Additional analyses may be conducted if deemed necessary for the interpretation of the data.

A more detailed description of these analyses will be provided in a separate analysis plan.

For summary analysis, any assay result below the lower limit of quantification (BLQ) will be considered as zero.

If dose-response or exposure-response modeling is warranted, it will be documented in a separate SAP.

### 7.9.2 Pharmacodynamic Analysis

Not applicable.

## 7.10 Other Outcomes

The pharmacogenomic data will not be included in the CSR.

For biomarker data, tables and listings will be prepared. Summary statistics will be done for C-reactive protein (CRP), Glucose-dependent insulinotropic polypeptide (GIP), and Peptide YY (PYY) by treatment group for overall for the actual values and change from baseline values.

## 7.11 Safety Analysis

All safety summaries will be based on the safety analysis set. Safety summaries will include descriptive statistics for values, changes, and incidence of events for each TAK-906 dose level and total TAK-906 combined (regardless of TAK-906 dosing level) and placebo.

### 7.11.1 Adverse Events

All adverse events will be coded using MedDRA latest version. All adverse events will be included in the data listings but only treatment emergent adverse events will be included in the summary tables.

A treatment-emergent adverse event (TEAE) will be defined as an AE or serious adverse event (SAE) that started or worsened after first study drug administration and within 30 days of last dose of study drug (AE onset date – date of last dose ≤30). AEs with missing onset dates will be summarized with TEAEs regardless of severity and relationship to study medication.

If AE start date is missing, it will be considered TEAE unless AE end date is prior to randomization date.

If a part of the AE start date or AE end date is missing, parts of randomization date will be assumed to derive the TEAE flag. For example, if only month is missing for ae start date, then randomization month will be used to derive TEAE flag.

The following summaries will be presented:

- Overview of TEAEs during the study - number and percentage of subjects, number of events.
- TEAEs by SOC and PT - number and percentage of subjects.
- TEAEs by PT - number and percentage of subjects.
- Intensity of TEAEs by SOC and PT - number and percentage of subjects.
- Relationship of TEAEs by SOC and PT - number and percentage of subjects.
- Drug-related TEAEs by SOC and PT - number and percentage of subjects.
- TEAEs Related to Study Medication by SOC and PT - number and percentage of subjects.
- TEAEs leading to study discontinuation by SOC and PT - number and percentage of subjects.
- Serious treatment-emergent AEs by SOC and PT - number and percentage of subjects, number of events.
- Treatment-emergent Adverse Events of Special Interest by SOC and PT - number and percentage of subjects, number of events.
- Most Frequent (>=5%) Treatment-emergent Non-serious AEs by SOC and PT - number and percentage of subjects, number of events.

- Most Frequent (>=5%) Treatment-emergent AEs by SOC and PT - number and percentage of subjects, number of events.
- Pretreatment Events by SOC and PT - number and percentage of subjects.

SOCs will be sorted in descending order of total number of subjects with the SOC among all the treatment groups. Within an SOC, adverse events will be sorted in descending order of total number of subjects with the preferred term among all the treatment groups.

In the high-level adverse event summary tables, TEAEs will be summarized regardless of intensity and relationship to study drug. Within each subject, multiple reports of events that map to a common MedDRA term will be counted only once.

At the adverse event level, the summary tables will present the number of subjects reporting each of these MedDRA events, ie, the number of subjects reporting 1 or more events that map to the given MedDRA term.

At the SOC level, the summary tables will present the number of subjects reporting 1 or more events that map to the given SOC. That is, the number of subjects reported at the SOC level will be less than or equal to the sum of the subject counts across all adverse events within that SOC.

In selected summaries (see above), adverse events will be summarized by the number of events reported in addition to the number and percentage of subjects with events.

For the summary of TEAEs by SOC, preferred term and maximum intensity, if a subject experiences more than 1 episode of a particular coded adverse event, the subject will be counted only once by the maximum intensity of the episode (preferred term). Similarly, if a subject has more than 1 adverse event within an SOC, the subject will be counted only once by the maximum intensity in that SOC.

TEAEs classified in the eCRF as possibly or probably related to the study medication will also be summarized by preferred term and SOC. If a subject experiences more than 1 episode of a particular coded adverse event, the subject will be counted only once by the most related report for the preferred term. Similarly, if a subject has more than 1 adverse event within an SOC, the subject will be counted only once by the most related report in that SOC. Adverse events with missing relationship will be classified as related to study drug.

All adverse events will be listed by treatment, study center, subject number and onset date of the adverse event. The listing will contain: subject identifier, age, sex, body weight, race, adverse event (preferred term and reported term), SOC, onset date, end date or whether the event was ongoing, duration, frequency, intensity (mild, moderate or severe), action taken concerning study drug, causality to study drug, the outcome, whether the adverse event was an SAE and whether the event was an adverse event of special interest (AESI).

Special listings for TEAEs leading to study discontinuation, SAEs, deaths, related AEs, and AE of special interest will also be presented.

### 7.11.2 Clinical Laboratory Evaluations

For each laboratory parameter, the following will be displayed for each scheduled time point (each visit and end of treatment) as applicable.

- Summary statistics (n, mean, SD, median, minimum, and maximum) by treatment group and overall for the actual values and change from Baseline values.
- Shift tables for the change from Baseline to each post-baseline time point will be presented. For these tables, each subject will be categorized as low, normal, or high for the baseline value, and low, normal, or high for each post-baseline time point, according to the central laboratory reference ranges. The number of subjects in each of the combinations of shifts will be presented.
- Markedly abnormal values (MAVs), as defined in Appendix B, will be summarized by treatment group and overall. The number and percentage of subjects with MAV values observed post-baseline in each of the applicable laboratory parameters will be presented.
- In addition to the MAV summary, for the following laboratory parameters, the number and percentages of subjects with the worst post-baseline values that fall in each of the listed categories will be summarized by treatment group:
  - Prolactin:
    - >ULN;
    - >2.5 × ULN;
    - >5 × ULN;
    - >10 × ULN.
  - Glucose:
    - <2.2 mmol/L;
    - <1.7 mmol/L;
    - >13.9 mmol/L;
    - ≥27.8 mmol/L.

  Liver Test Parameters

  - ALT:
    - >3 × ULN;
    - >3 – ≤5 × ULN;
    - >5 – ≤8 × ULN;
    - >8 – ≤20 × ULN;

- >20.0 × ULN.

– AST:
  - >3 × ULN;
  - >3 – ≤5 × ULN;
  - >5 – ≤8 × ULN;
  - >8 – ≤20 × ULN;
  - >20 × ULN.

– ALT or AST:
  - ALT >3 × ULN or AST >3 × ULN;
  - (ALT >3 × ULN and ALT ≤5 × ULN) or (AST >3 × ULN and AST ≤5 × ULN);
  - (ALT >5 × ULN and ALT ≤8 × ULN) or (AST >5 × ULN and AST ≤8 × ULN);
  - (ALT >8 × ULN and ALT ≤20 × ULN) or (AST >8 × ULN and AST ≤20 × ULN);
  - ALT >20 × ULN or AST >20 × ULN.

– ALT/AST and Total Bilirubin:
  - ALT >3 × ULN AND TBILI >2 × ULN;
  - AST >3 × ULN AND TBILI >2 × ULN;
  - (ALT >3 × ULN or AST >3 × ULN) AND TBILI >2 × ULN.

A listing of all laboratory data will be provided. Laboratory data outside of the normal reference range will be flagged on the listing along with values meeting MAV criteria.

Summaries and listings of laboratory data will be presented in System International (SI) unit.

### 7.11.3 Vital Signs and Weight

Vital signs and weight at scheduled visits and their changes from Baseline will be summarized for each treatment group and overall using descriptive statistics by visit and end of treatment. The number and percentage of subjects with at least one post-baseline abnormal vital sign value that falls in the each of the category defined in Table 7.c will be tabulated for each variable across all visits.

Bot plots will be provided for vial sign parameters for each treatment group by visit and end of treatment visit.

**Table 7.c Categories of Abnormal Vital Signs Values**

| Parameter | Low Abnormal | High Abnormal |
|---|---|---|
| Pulse rate (HR) | <50 bpm | >120 bpm |
| | Decrease from baseline >20 bpm | Increase from baseline >20 bpm |
| | <50 bpm **AND** decrease from baseline >20 bpm | >120 bpm **AND** increase from baseline >20 bpm |
| SBP | <90 mmHg | >160 mmHg |
| | >20 mmHg decrease from baseline | >20 mmHg increase from baseline |
| | <90 mmHg **AND** >20 mmHg decrease from baseline | >160 mmHg **AND** >20 mmHg increase from baseline |
| DBP | <60 mmHg | >90 mmHg |
| | >10 mmHg decrease from baseline | >10 mmHg increase from baseline |
| | <60 mmHg **AND** >10 mmHg decrease from baseline | >90 mmHg **AND** >10 mmHg increase from baseline |
| Body Temperature | < 35.0 $^{o}$C | >38.3 $^{o}$C |

A listing of all vital signs and weight data will be provided. Values meeting the MAV criteria will be flagged. The criteria for identification of vital signs MAVs are given in Appendix C.

### 7.11.4 12-Lead ECGs

ECG variables at scheduled visits and their changes from Baseline will be summarized for each treatment group and overall using descriptive statistics by study visit and end of treatment. A shift table for the investigator's ECG interpretation will provide the number of subjects in each of the appropriate categories (Normal, Abnormal but not clinically significant, or Abnormal and clinically significant) at the scheduled visit relative to the Baseline status. The number and percentage of subjects with at least one post-baseline abnormal ECG value that falls in each of the category defined in Table 7.d will be tabulated for each variable across all visits.

Box plots will be provided for ECG parameters for each treatment group by visit and end of treatment visit.

**Table 7.d Categories of Abnormal ECG Values**

| Parameter | Low Abnormal | High Abnormal |
|---|---|---|
| Heart rate (HR) | <50 bpm | >120 bpm |
| | Decrease from baseline >20 bpm | Increase from baseline >20 bpm |
| | <50 bpm **AND** decrease from baseline >20 bpm | >120 bpm **AND** increase from baseline >20 bpm |
| PR | <120 milliseconds | >200 milliseconds |
| | | >200 milliseconds **AND** increase from baseline >25% |
| QRS | <60 milliseconds | >120 milliseconds |
| | - | >120 milliseconds **AND** increase from baseline >25% |
| QTcF | - | >450 and ≤480 milliseconds |
| | - | >480 and ≤500 milliseconds |
| | - | >500 milliseconds |
| | - | Change from baseline of >30 and ≤60 milliseconds |
| | - | Change from baseline of >60 milliseconds |
| | - | >450 milliseconds **AND** >30 milliseconds change from baseline |

A listing of all ECG data will be provided. Values meeting the MAV criteria will be flagged. The criteria for identification of ECG MAVs are given in Appendix D.

### 7.11.5 Other Observations Related to Safety

For female subjects, pregnancy test results will be listed by treatment, study center and subject number.

## 7.12 Interim Analysis

Based on Protocol Amendment 8, interim analysis has been canceled.

## 7.13 Multiplicity Adjustments

**There are no multiplicity adjustment proposed for this study.**

## 7.14 Changes in the Statistical Analysis Plan

| Version | Description | Changes | Rationale |
|---|---|---|---|
| 1 | Original Document based on Protocol Amendment 6, 8 August 2019 | | |

| Version | Description | Changes | Rationale |
|---|---|---|---|
| 2 | Based on Protocol Amendment 8, 13 July 2020 | | |
| | | Interim Analysis was removed.<br>2-sided analysis was modified to 1-sided analysis at alpha 0.05.<br>Data Collection was modified in Protocol Amendment 8, analysis will still be conducted for subjects who entered in a previous version of the protocol. Ex) The data collection GEBT has changed in Protocol Amendment 8.<br>PGI-S Scores version 1was replace with version 2.<br>PGI-I Scores was changed to PGI-C Scores.<br>Added an assessment on COVID-19 impacted subjects. | Due to modifications in Protocol Amendment 8, the following were implemented. |
| | | Per-Protocol Definition Clarified. | To narrow down the factors important to objective of study. |
| | | Clarified how to handle AE date missing in deriving TEAE flag. | There was no clarification on missing dates in the previous SAP. |
| | | ANOVA and Cochran-Mantel-Haenszel for baseline variables are removed. | Descriptive statistics are sufficient for baseline variables. |
| | | AUC analysis removed. | To simplify analysis. |
| | | Sensitivity Analysis modified. | To provide robust analysis amongst subjects with rescue medication. |
| | Changes from the Protocol Version 8.0 | For both continuous and categorical analysis, center will not be used as a fixed factor. The mixed models will have all the factors from the protocol except the center variable. | Insufficient number of subjects enrolled in a site may lead to small sample size in a strata. It is assumed that center to center variability in subjects are similar since following one protocol. |
| | | Safety Analysis set updated to be at least 1 dose of study drug | To evaluate safety in all subjects that were exposed to any amount of study drug. |
| 3 | | Updated drug compliance calculation. | Updated drug compliance calculation to address the missing bottles. |
| | | Set the recorded vomiting frequency that is clinically Implausible as missing. | Recorded vomiting frequency above 60 per day is considered clinically implausible and will be set to missing, and summarized as recorded vomiting frequency. |

| Version | Description | Changes | Rationale |
|---|---|---|---|
| | | Summary tables added for impact due to COVID-19, CGI-I, CGI-S, PGI-I and PGI-C. | Summary analysis are added to help interpret data in addition to by-subject listing. |
| | | PK summary table is added.<br>Biomarker summary table is added. | Summary analysis are added to help interpret data in addition to by-subject listing. |
| | | Clarifications added for<br>50% reduction from baseline in the weekly ANMS GCSI-DD<br>>=20% reduction in last 6 consecutive weeks,<br>Pattern Mixture Sensitivity Analysis | Clarifications added to help implementation. |
| | | Safety Analysis Set definition updated to handle if subject received more than one treatment | To have conservative approach for safety analysis when subject received more than one treatment. |
| | | Removed 'at least 1 valid postbaseline value' in the definition for FAS. | Address FDA's comments. |
| | | Sensitivity analysis 1 and 2 modified to incorporate the timing of when the rescue medication was used. | Address FDA's comments. |
| | | Clarifications added for handling irregular data<br>[1] Analyses only include data evaluated by trained and qualified evaluators.<br>[2] Clarified to how to handle duplicate data entry, data entered not on the same day as planned. | [1] Exclude data evaluated by non-trained and non-qualified evaluators to avoid misleading information.<br>[2] Use algorithms to derive robust dataset for analysis. |
| | | SF-12 MCS and PCS derivations will be based on PRO Core® software. | Use commercial software to derive SF-12 MCS and PCS since the algorithm to the norm-based scoring is proprietary. |



## 8.0 REFERENCES

1. A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Phase 2b Study to Evaluate the Efficacy and Safety of Twice-Daily Oral Administration of a Peripherally Acting Dopamine Receptor $D_2/D_3$ Antagonist, TAK-906 for the Treatment of Adult Subjects With Symptomatic Idiopathic or Diabetic Gastroparesis, Millennium Pharmaceuticals, Inc., a wholly owned subsidiary of Takeda Pharmaceutical Company Limited, Protocol No. TAK-906-2002, Amendment 08, dated 13 July, 2020.

2. Revicki DA, Camilleri M, Kuo B, et al. Evaluating symptom outcomes in gastroparesis clinical trials: validity and responsiveness of the Gastroparesis Cardinal Symptom Index-Daily Diary (GCSI-DD). Neurogastroenterol Motil 2012;24:456-463.
3. Rentz AM, Kahrilas P, Stanghellini V, et al. Development and psychometric evaluation of the patient assessment of upper gastrointestinal symptom severity index (PAGI-SYM) in patients with upper gastrointestinal disorders. Qual Life Res 2004;13:1737-1749.
4. Szarka L, Camilleri M, Vella A, et al. A Stable Isotope Breath Test With a Standard Meal for Abnormal Gastric Emptying of Solids in the Clinic and in Research. Clinical Gastroenterology and Hepatology 2008;6:635–643
5. Kieser M, Friede T. Simple procedures for blinded sample size adjustment that do not affect the type I error rate. Statistics in Medicine. 2003; 22: 3571-3581.
6. Kiernan K, Tao J, Gibbs P. Tips and strategies for mixed modeling with SAS/Stat procedures. SAS global forum 2012.
7. Lu, K, Mehrotra, DV. Specification of covariance structure in longitudinal data analysis for randomized clinical trials. Statistics in Medicine. 2010; 29: 474–488.
8. Ware JE, Kosinski M, Keller SD. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. Boston, MA. The Health Institute, New England Medical Center, Second Edition, 1995.
9. Bohdana Ratitch, Michael O'Kelly. Implementation of Pattern-Mixture Models using Standard SAS/STAT Procedures. Available at link: https://www.pharmasug.org/proceedings/2011/SP/PharmaSUG-2011-SP04.pdf. (accessed 10August2020)Maruish, M. E. (Ed.). (2012). User's manual for the SF-12v2 Health Survey (3rd ed.). Lincoln, RI: QualityMetric Incorporated.
10. Lingling Li. SAS® V9.4 MNAR statement for multiple imputations for missing not at random in longitudinal clinical trials. Available at link: https://www.pharmasug.org/proceedings/2019/ST/PharmaSUG-2019-ST-103.pdf (Accessed 10August2020).
11. Parkman, Henry, et al. User Manuel for the ANMS GCSI-DD (2018) Available at link: https://www.fda.gov/media/125038/download (Accessed 29July2021)

## Appendix A Numeric Conversions of PRO data

| Parameter | Numeric Conversions |
|---|---|
| Composite Score<br>Total Symptom Score<br>Core Symptom Score | For nausea, early satiety, postprandial fullness, upper abdominal pain, bloating severity and overall gastroparesis severity, the responses will be coded into following numeric values: 0 = None, 1 = Mild, 2 = Moderate, 3 = Severe, 4 = Very Severe<br><br>The vomiting frequency will be scored from 0 to 4 (where 0 = no vomiting and 4 = four or more episodes of vomiting) |
| PAGI-SYM_TOT<br>PAGI-SYM_NAVM<br>PAGI-SYM_PFES<br>PAGI-SYM_BL<br>PAGI-SYM_UAP<br>PAGI-SYM_LAP<br>PAGI-SYM_HBRE | The PAGI-SYM contains 20 items grouped into 6 subscale scores. A 6-point Likert response scale, ranging from 0 (none) to 5 (very severe), is used for rating the severity of each symptom. |
| CSSR_TOT | CSSR total score will be an average of all the items, in which each item ranges from 0 (none) to 5 (very severe). |
| PAGI-QoL | PAGI-QoL Responses will be coded in the numeric values; 0= None of the time, 1= A Little of the time, 2= Some of the time, 3= A good bit of the time, 4= Most of the time, 5= All of the time<br>30 items of PAGI-QOL cover five domains/subscales:<br>• Daily Activities<br>• Clothing, Diet and Food Habits<br>• Relationship<br>• Psychological Well-Being<br>• Distress<br>5 dimensions: Daily Activities (10 items related to avoiding or having difficulties with daily activities), Clothing (2 items, one related to feeling constricted and one related to the frustration felt about not being able to dress as wanted), Diet and Food Habits (7 items related to restrictions made and induced frustration), Relationship (3 items describing the impact of the disease on relationships with their partner, relatives, and friends), and Psychological Well-being and Distress (8 items describing disease impact on feelings or emotional state) |

## Appendix B Criteria for Markedly Abnormal Values for Lab Parameters

| Parameter | Lower Criteria | Upper Criteria |
|---|---|---|
| Prolactin | | >5 x ULN |
| Albumin | <20 g/L | |
| Total protein | <0.8 x LLN | >1.2 x ULN |
| Creatinine | | >1.5 x ULN |
| | | |
| Blood urea nitrogen | | >1.5 x ULN |
| Creatine kinase (CPK) | | >5 x ULN<br>>10 × ULN |
| Potassium | <3.0 mmol/L | >5.5 mmol/L |
| Sodium | <130 mmol/L | >155 mmol/L |
| Calcium | <2.0 mmol/L | >2.9 mmol/L |
| Chloride | <0.8 x LLN | >1.2 x ULN |
| Magnesium | <0.4 mmol/L | >1.23 mmol/L |
| Phosphate | <0.6 mmol/L | |
| Total cholesterol | | >10.34 mmol/L |
| Triglycerides | | >5.7 mmol/L |
| High-density lipoprotein cholesterol | <LLN | |
| Low-density lipoprotein cholesterol | | >ULN |
| Glucose | <2.2 mmol/L | >13.9 mmol/L |
| Uric Acid | | >1.2 x ULN |
| ALT | | >3 x ULN |
| AST | | >3 x ULN |
| Total Bilirubin | - | >2 x ULN |
| GGT | - | >2.5 x ULN |
| Alkaline Phosphatase | - | >2.5 x ULN |
| Direct bilirubin | - | >1.5 x ULN |
| HbA1c | - | >1.2 x ULN |
| Red blood cells | <0.8 x LLN | >1.2 x ULN |
| White blood cells | <2.0 x $10^9$/L | >100 x $10^9$/L |
| Neutrophils | <1.0 x $10^9$/L | |
| Lymphocytes | < 0.5 x $10^9$/L | >20 x $10^9$/L |
| Hemoglobin | <80 g/L | >40 g/L above ULN |
| Hematocrit | <0.8 x LLN | >1.2 x ULN |
| Platelets | <75.0 x $10^9$/L | >600 x $10^9$/L |
| Mean corpuscular hemoglobin | <0.8 x LLN | >1.2 x ULN |
| PT | | >2.5 x ULN |
| INR | | >2.5 x ULN |

## Appendix C Criteria for Markedly Abnormal Values for Vital Signs

| Parameter | Unit | Lower Criteria | Upper Criteria |
|---|---|---|---|
| Heart Rate | Bpm | <50 or decrease from baseline >20 | >120 or increase from baseline of >20 |
| Systolic blood pressure | mm Hg | <90 or decrease from baseline >20 | >160 or increase from baseline of >20 |
| Diastolic blood pressure | mm Hg | <60 or decrease from baseline of >10 | >90 or increase from baseline of >10 |
| Body temperature | $^{o}$C | <35.0 | >38.3 |
| | $^{o}$F | <95 | >100.9 |

## Appendix D Criteria for Markedly Abnormal Values for the 12-Lead ECG Parameters

| Parameter | Lower Criteria | Upper Criteria |
|---|---|---|
| Heart rate | <50 bpm or decrease from baseline of >20 | >120 bpm or increase from baseline of >20 |
| PR | <120 milliseconds | >200 milliseconds |
| QTcF Interval | | >500 milliseconds or<br>>30 milliseconds increase from baseline and >450 milliseconds |
| QRS | <60 milliseconds | >120 milliseconds |

## Appendix E Details for Pattern Mixture

**Step 1:** use Proc MI's Monte Carlo Markov Chain (MCMC) statement, to produce monotone missing data patterns to impute the outcome variable at consecutive visits in a sequential manner.

- Set the imputation model including the treatment, stratum, and weekly composite scores at baseline, and visit 1 through 12 in the VAR statement, where the Minimum and Maximum for weekly composite scores at baseline, and visit 1 through 12 are set to zero and four, respectively.
- Set random seed and use MCMC statement with single chain and impute = monotone to create the 100 copies (numbered by _Imputation_) of dataset with monotone missing patterns and set MINMAXITER as 10000 to make sure all imputed values within the specified values of Minimum and Maximum.

**Step 2:** Assemble the dataset from step 1, use monotone regression and MNAR statement to impute missing weekly composite scores, by _Imputation, where the imputation model is derived by data from control group (ie, TRT01PN=1) only. If visit t is complete for all subjects, then MNAR statement would start from visit t+1.

- Set the imputation model including the stratum, and weekly composite scores at baseline, and visit 1 through 12 in the Var statement with treatment and stratum in Class statement, where the Minimum and Maximum for weekly composite scores at baseline, and visit 1 through 12 are set to zero and four, respectively.
- Set random seed and use monotone regression and MNAR statement with MODELOBS= (TRT01PN='1') specifying that the control group is used to derive the imputation model to impute missing weekly composite scores in the order listed in the VAR statement in step 1.

**Step 3:** Assemble the imputed dataset from step 2 to derive the primary endpoint, analyze the primary endpoint using MMRM with same fixed effects and covariates in section 7.8.2.1 and with unstructured covariance matrix by _Imputation_ to obtain the lsmeans difference between treatments (TAK-906 vs Placebo) at each visit.

**Step 4:** the results (lsmeans estimates and standard errors) from step 3 will be combined by Rubin's rules using Proc MIANALYZE by visit to estimate treatment differences and associated 90%CIs at each visit. The 1-sided p-value for testing treatment differences (TAK-906 vs Placebo) in primary endpoint at each visit will be based on the t-test with degrees of freedom from the MMRM in step 3.

## Appendix F Convention for Missing dates for Adverse Events/Concomitant Medication/Procedures

Conventions for Missing/Partial Dates for Adverse Event/Concomitant Medication/Procedures

The start date that is completely or partially missing will be imputed as follows:

- If month and year are known but day is missing
  - If month and year are the same as month and year of the 1st dose date, the day of the 1st dose date will be used to impute the missing day.
  - If month and year are prior to the month and year of the 1st dose date, the last day of the month will be used to impute the missing day.
  - If month and year are after the month and year of the 1st dose date, the 1st day of the month will be used to impute the missing day.
- If year is known, but both day and month are missing
  - If the year is same as year of the 1st dose date, the month and day of the 1st dose date will be used to impute the missing month and day, respectively.
  - If the year is prior to the year of the 1st dose date, December 31st of the year will be used to impute the missing month and day, respectively.
  - If the year is after the year of the 1st dose date, January 1st of the year will be used to impute the missing month and day, respectively.
- If all (day, month, year) are missing, the 1st dose date will be used to impute the missing year, month, and day, respectively.

Imputing missing start date is mandatory. After imputation, all imputed dates are checked against the stop dates to ensure that start dates do not occur after stop dates. If an imputed start date occurs after the stop date, then change the imputed start date to be the same as the stop date.

The stop dates that are completely or partially missing will be imputed as follows:

- If the AE is "ongoing", no imputation is necessary.
- If month and year are known but day is missing, the last day of the month will be used to impute the missing day.
- If year is known, but both day and month are missing
  - December 31st of the year will be used to impute the missing month and day, respectively.
- If all (day, month, year) are missing, the event will be considered as ongoing.

Imputing missing stop date is not mandatory if event is considered as ongoing. However if it is to be done, the rules are outlined above. If subject dies, then use death date for the stop date. After imputation, all imputed dates are checked against start dates to ensure that stop dates do not occur before start dates. If an imputed stop date occurs prior to the start date, then change the imputed stop date to be the same as the start date.

## ELECTRONIC SIGNATURES

| Signed by | Meaning of Signature | Server Date (dd-MMM-yyyy HH:mm 'UTC') |
|---|---|---|
| PPD | Biostatistics Approval | 09-Aug-2021 12:46 UTC |