Psychometric Validation Study of the ANMS GCSI-DD

Statistical Analysis Plan

EVA-20216-01 | May 14, 2018 | Version 7.0

Prepared For:
Jorge Puelles, PharmD, MSc, MBA
Director, Global Outcome Research
Takeda Development Center Europe

Prepared By:
Sara Gleeson, MPH
Dennis Revicki, PhD
Rebecca Speck, PhD, MPH

Project Contact:
Rebecca Speck | Tel +1.301.266.1439 | rebecca.speck@evidera.com

Table of Contents

List of Tables

1.0 Background

The American Neurogastroenterology and Motility Society Gastroparesis Cardinal Symptom Index-Daily Diary (ANMS GCSI-DD) is the first instrument to provide valid, reliable, and standardized data on the patient-reported symptoms of gastroparesis and the effect of treatment, and is intended for use as a primary or secondary endpoint in clinical trials. Five core symptoms are monitored in the ANMS GCSI-DD: nausea, vomiting, early satiety, postprandial fullness, and upper abdominal pain. To improve patient recall in reporting symptoms, the ANMS GCSI-DD is completed as a daily diary (i.e., 24-hour recall).

The ANMS GCSI-DD was originally developed based on concept elicitation focus groups and qualitative interviews, and was refined through a process of cognitive interviews—all with representative patients from the target gastroparesis population. Based on the requirements of the Food and Drug Administration (FDA) patient-reported outcome (PRO) guidance (2009), the recent FDA guidance on gastroparesis (2015), and FDA feedback to date during the qualification guidance of the ANMS GCSI-DD, a qualitative interview study was recently conducted by Evidera to generate additional support for the content validity of the revised ANMS GCSI-DD. The qualitative research phase was conducted in a manner that was conclusive with respect to concept elicitation, and demonstrated that sufficient cognitive interviews were conducted to ensure comprehension of the instructions, items, and response scales.

As Takeda's TAK-906 clinical trial program will include a finalized version of the ANMS GCSI-DD, additional psychometric evaluation evidence is needed to further evaluate and confirm the measurement properties of the daily diary prior to implementation in Takeda's Phase 2 and subsequent Phase 3 studies in gastroparesis. Specifically, the data and analyses planned herein from Evidera's standalone psychometric study and Takeda's TAK-906 Phase 2a trial aim to provide sufficient supportive evidence of the unidimensionality, item performance, reliability (i.e., internal consistency, test-retest) and validity (i.e., concurrent, known groups) of the ANMS GCSI-DD.

2.0 Analysis Objective

This statistical analysis plan (SAP) is designed to assess the following characteristics of the ANMS GCSI-DD:

- Unidimensionality and item performance
- Reliability (internal consistency and test-retest)
- Validity (concurrent and known groups)
- Exploration of responsiveness and responder definition threshold

3.0 Overview of Study Design

Data from two study populations will be used for the analyses described in this SAP, either alone or in combination. The two study populations are described in this section.

3.1 Overview of Psychometric Validation Study

This psychometric validation study (EVA-20216-01) was a longitudinal, four-week, observational study consisting of two study visits (Baseline and Week 4) and at-home diary completion for four weeks. There were no study drugs or invasive procedures administered as part of this study. Study patients were identified through medical chart review or during a regularly scheduled office visit.

A total of approximately 70 participants completed one in-person study visit at Baseline (Visit 1). Following Visit 1, participants were asked to complete the ANMS GCSI-DD and Bloating Severity Item on a daily basis for four weeks at home. Participants returned to the clinic for a Week 4 visit (Visit 2), approximately 30 days after the Baseline visit, for follow-up PRO and clinician assessments. Site staff checked completed questionnaires for any missing data at Visit 2. Site staff were responsible for making reminder calls prior to both visits. A schematic of the study is shown in Table A.

Table A. Schematic of Study Design for the EVA-20216-01 Psychometric Validation Study

Measure	Completed at Screening Prior to Baseline	Completed at Baseline (Visit 1) 0	Completed Daily at Home Between Visit 1 and Visit 2 28 Days	Completed at Week 4 (Visit 2) 28 ± 7 Days
Inclusion/Exclusion Form[1]	X			
Sociodemographic Form[2]		x		
PAGI-SYM[2]		x		x
PGI-S[3]		x		x
CSSR[3]		x		x
PROMIS Global Scale[2]		x		x
CGI-S[3]		x		x
Clinical Form[3]		x		
ANMS GCSI-DD[2]			x	
Patient-rated Bloating Severity Item[2]			x	
PGI-I[2]				x
CGI-I[3]				x

[1] Site staff-completed

[2] Patient-completed

[3] Clinician-completed

PAGI-SYM = Patient Assessment of Upper Gastrointestinal Disorders-Symptom Severity Index; PGI-S = Patient Global Impression of Severity; CSSR = Clinical Symptom Severity Rating; PROMIS = Patient-Reported Outcomes Measurement Information System; CGI-S = Clinician Global Impression-Severity; ANMS GCSI-DD = American Neurogastroenterology and Motility Society Gastroparesis Cardinal Symptom Index-Daily Diary; CGI-I = Clinician Global Impression-Improvement; PGI-I = Patient Global Impression of Improvement

3.1.1 Psychometric Validation Study Inclusion Criteria

Participants had to meet **all** of the following criteria to be considered for enrollment into this study:

1. Adult men and women aged 18 to 85 years, inclusive;
2. BMI ≥18 and ≤35kg/m^2;
3. Have experienced symptoms of diabetic or idiopathic gastroparesis (e.g., postprandial fullness, nausea, vomiting, upper abdominal pain and early satiety) (at least intermittently) for at least 6 months prior to screening as assessed by a physician;
4. Diagnosis of gastroparesis with documented delay in gastric emptying according to local clinical criteria;
5. Gastroparesis from either idiopathic or diabetic etiologies;
6. Available to attend two in-person study visits;
7. Able and willing to complete a diary for approximately four weeks at home;
8. Able and willing to provide written informed consent prior to participation in the study;
9. Able to speak and read English; and

Special inclusion for patients with diabetic gastroparesis:

10. Has diabetes mellitus (DM) with HBA1c <11%.

3.1.2 Psychometric Validation Study Exclusion Criteria

Participants were not eligible for inclusion in this study if **any** of the following criteria apply:

1. Prior history of gastric surgery, including but not limited to gastrectomy, gastric bypass, gastric banding, pyloroplasty, vagotomy, or fundoplication, which has manipulated the natural anatomy of the stomach;
2. Known secondary causes of gastroparesis including but not limited to Parkinson's Disease, cancer, viral illness, or connective tissue diseases;
3. History of intrapyloric botulinum toxin injection within 3 months of Screening or currently has functioning implantable electric stimulator;
4. Have a history of clinically significant endocrine, GI (including motility disorder, intestinal obstruction, inflammatory bowel disease), cardiovascular, hematological, hepatic, immunological, renal, respiratory, genitourinary, or major neurological (including stroke and chronic seizures) abnormalities or diseases;
5. Predominant symptom is epigastric pain, diffuse abdominal pain, or pain associated with bowel movement;
6. Has a history of anorexia nervosa or bulimia;
7. Previous history of bezoars; or

8. Patient has any clinically relevant condition that in the opinion of the investigator/coordinator, would interfere with completing the study including, but not limited to visual problems, severe mental illness or cognitive impairment.

3.2 Overview of TAK-906 Phase 2a Clinical Trial

TAK-906-1002 was a two-part, randomized, double-blind and open-label, placebo and active-comparator controlled trial to evaluate the safety, pharmacokinetics, and pharmacodynamics for TAK-906 in subjects with diabetes mellitus and gastroparesis or idiopathic gastroparesis. Part 1 data will be used in the analyses described herein.

In Part 1, a total of approximately 48 subjects were randomized into one of three active treatment arms (TAK-906 maleate 5 mg twice daily [BID], 25 mg BID, or 100 mg BID) or a placebo arm in a double-dummy manner for nine consecutive days. Subject randomization was stratified by DG versus IG. Key safety and tolerability endpoints will be assessed during Part 1 through physical exam, vital signs, 12-lead electrocardiogram (ECG), clinical laboratory assessments, and collection of serious and non-serious adverse events (AEs). After completion of the trial (or following subject withdrawal), all subjects will return for a follow-up visit 10-14 days after their last dose of the study medication. A schematic of the relevant study assessments is shown in Table B.

Table B. Schematic of Study Design for TAK-906-1002

Measure	Screening Day -28 to Day -3	-2	-1	Pre-dose Day 1	1	2	3	4	5	6	7	8	9	Early Termination	Follow-up[1]
Medical history and demographics	x														
Study drug administration					x	x	x	x	x	x	X	x	x		
Full physical examination	x			x								x		x	x
Vital signs	x			x										x	x
GEBT	x		x		x						X				
SmartPill ingestion			x								X				
Nausea subscore of ANMS GCSI-DD	x														
ANMS CGSI-DD	x	x	x		x	x	x	x	x	x	X	x	x	x	
PAGI-SYM				x									x		
CSSR				x									x		

[1] Follow-up will occur approximately 14 days after the last dose of trial drug received.

GEBT = gastric emptying breath test; ANMS GCSI-DD = American Neurogastroenterology and Motility Society Gastroparesis Cardinal Symptom Index-Daily Diary; PAGI-SYM = Patient Assessment of Upper Gastrointestinal Disorders-Symptom Severity Index; CSSR = Clinical Symptom Severity Rating

3.2.1 TAK-906 Phase 2a Inclusion Criteria

Participants had to meet **all** of the following criteria to be considered for enrollment into this study:

1. Understand the study procedures and agree to participate by providing written informed consent;
2. Be willing and able to comply with all study procedures and restrictions;
3. Has a documented diagnosis of diabetic gastroparesis or idiopathic gastroparesis;
4. Be man or a woman aged 18 to 75 years, inclusive, at the Screening Visit;
5. Have a body mass index (BMI) ≥18 and ≤40 (kg/m²) at the Screening Visit;
6. Be a nonsmoker who has not used tobacco- or nicotine-containing products (e.g., nicotine patch) for at least six months prior to trial drug administration of the initial dose of trial drug/invasive procedure;
7. Be judged to be in good health by the investigator, based on clinical evaluations including laboratory safety tests, medical history, physical examination, 12-lead ECG, and vital sign measurements performed at the Screening Visit and prior to administration of the initial dose of trial drug/invasive procedure;
8. Has a QT interval with Fridericia correction method (QTcF) interval <450 mSec (men) <475 mSec (women);
9. Has a serum prolactin <2× upper limit of normal (ULN) at screening;
10. Has an alanine aminotransferase (ALT) and aspartate aminotransferase (AST) <1.5×ULN; bilirubin ≤1.5×ULN (except Gilbert syndrome);
11. Meet the trial birth control requirements (not shown);
12. Has symptoms for GP (i.e., chronic postprandial fullness, abdominal pain, postprandial nausea, vomiting, loss of appetite and/or early satiety) the past three months;
13. Has documented slow GE, with delayed GE by GEBT at Screening defined as ≥80th percentile (Note: If a subject has had a documented scintigraphy or GEBT within the last 12 months that confirms the diagnosis of delayed GE, a screening GEBT would not be required);
14. Has nausea subscale (of ANMS-GCSI-DD) symptom score ≥2 at least three of seven days during Screening; and
15. Has HBA1c <10% (for diabetes mellitus only). (Note: Given the biological variability of glycemic parameters, subjects with a value that does not meet the above criteria, but is within 0.2% HBA1c of the qualifying range may, at the discretion of the investigator, have a repeat determination performed and use as a qualifying parameter in lieu of the original value.)

3.2.2 TAK-906 Phase 2a Exclusion Criteria

Participants were not eligible for inclusion in this study if **any** of the following criteria apply:

1. Subjects who have a history of clinically significant endocrine, GI (including motility disorder and intestinal obstruction), cardiovascular, hematological, hepatic, immunological, renal, respiratory,

genitourinary, or major neurological (including stroke and chronic seizures) abnormalities or diseases will be excluded from the trial;

2. Has participated in another investigational trial within four weeks prior to the pretrial (Screening) visit (the four-week window will be derived from the date of the last trial procedure and/or AE related to the trial procedure in the previous trial to the pretrial/Screening Visit of the current trial);
3. Is an employee or immediate family member (e.g., spouse, parent, child, sibling) of the clinical site or of the sponsor;
4. Has acute severe gastroenteritis and pronounced dehydration in the past 48 hours prior to Screening, gastric pacemaker, chronic parenteral feeding or persistent severe vomiting;
5. Has a known disturbance of small intestinal absorption, exocrine pancreatic function, liver metabolism, and pulmonary function;
6. Has a history of anorexia nervosa or bulimia;
7. Has a history of additional risk factors for Torsade de Pointes (e.g., heart failure, hypokalemia, family history of Long QT Syndrome), evidence of cardiac autonomic neuropathy (e.g., lack of RR variation upon deep breathing);
8. Previous history of bezoars (the presence of retained liquid, bile, or small amounts of poorly organized food residue is permitted);
9. Difficulty swallowing solid food or pills;
10. Prior surgery involving the luminal GI tract (cholecystectomy, appendectomy, and hysterectomy are permitted if performed >3 months prior to SmartPill test);
11. Any abdominal or pelvic surgery within the past three months;
12. Known or history of inflammatory bowel disease;
13. Has active diverticulitis, diverticular stricture, and other intestinal strictures;
14. Has active ongoing cancer. In addition, subjects with a history of cancer who have received treatment within the last 5 years are not eligible to screen.
15. Has a history of significant multiple and/or severe allergies (e.g., food, drug, latex allergy) or has had an anaphylactic reaction or significant intolerability to prescription or nonprescription drugs or food;
16. Has a hypersensitivity to Spirulina, egg, milk, bread, or jam, or wheat allergens;
17. Has a positive alcohol or drug screen;
18. Lactating or pregnant women as determined by positive HCG test at Screening and within 24 hours of the first dose;
19. Is breastfeeding or has breastfed in approximately the last 6 months.
20. Is positive for hepatitis B surface antigen (HBsAg), hepatitis C antibodies, or HIV (confirmatory testing is allowed; most sensitive test should take precedence);

21. Had major surgery, donated or lost 1 unit of blood (approximately 500 mL) within four weeks prior to the pretrial (screening) visit;
22. Is unable and/or unwilling to refrain from or anticipates the use of any medication, including prescription and nonprescription drugs or herbal remedies beginning approximately at least 10 days prior to assessment of GP on Day -1, throughout the trial (including washout intervals between treatment periods), until the post-trial visit (there may be certain medications that are permitted);
23. Chronic daily use of nonsteroidal anti-inflammatory drugs (ibuprofen, naproxen, etc.);
24. Uses medications that prolong the QT/QTc interval;
25. Has a history of alcohol consumption exceeding two standard drinks per day on average (one glass is approximately equivalent to: beer [354 mL/12 ounces], wine [118 mL/4 ounces], or distilled spirits [29.5 mL/1 ounce] per day);
26. Consumes excessive amounts, defined as greater than six servings (one serving is approximately equivalent to 120 mg of caffeine) of coffee, tea, cola, energy drinks, or other caffeinated beverages per day;
27. Has a substance-abuse disorder;
28. Has a consistent fasting glucose of approximately ≥270 mg/dL (14.99 mmol/L) during any visit up to and including the randomization visit (Period 1 Day 1 predose);
29. Has had diabetic ketoacidosis (within the prior four weeks);
30. Use of cardiac medical devices such as pacemakers and defibrillators (bladder stimulators, spinal stimulators, medication infusion devices, insulin pumps, continuous glucose monitors) are permitted
31. Subject is unable or unwilling to comply with the study rules, regulations, and procedures.

4.0 Description of Relevant Measures

Table C illustrates the timing of the below measures in the EVA-20216-01 and TAK-906-1002 studies.

4.1 ANMS GCSI-DD

The ANMS GCSDI-DD is a daily diary that consists of six questions based on the original Gastroparesis Cardinal Symptom Index, with revisions based on interactions with the FDA. The ANMS GCSI-DD includes a 24-hour recall period, and is intended for daily administration. Items included in the ANMS GCSI-DD include severity assessments of nausea, early satiety, post-prandial fullness, and upper abdominal pain, vomiting frequency, and an overall assessment of gastroparesis symptom severity. The severity of gastroparesis symptoms is assessed using a five-point Likert scale (None, Mild, Moderate, Severe, Very Severe). The scores range from 0 (None) to 4 (Very Severe). Vomiting is assessed based on frequency (episodes of vomiting) over the past 24 hours.

4.1.1 ANMS GCSI-DD Endpoint Score for the TAK-906 Clinical Trial Program

The ANMS GCSI-DD symptoms that are of interest as an endpoint for the TAK-906 clinical trial program are nausea, early satiety, postprandial fullness, and upper abdominal pain. The ANMS GCSI-DD endpoint score is generated by summing the scores on the four symptom items and then dividing by 4 (i.e., the number of items within the endpoint score).

** This measure was used in EVA-20216-01 and TAK-906-1002.*

4.1.2 ANMS GCSI-DD Core Symptom Composite Score

The ANMS GCSI-DD core symptom composite score is generated by summing the scores on each of the five symptom items (nausea, vomiting, early satiety, postprandial fullness, and upper abdominal pain) and then dividing by 5 (i.e., the number of items within the core symptom composite score).

** This measure was used in EVA-20216-01 and TAK-906-1002.*

4.1.3 Patient-reported Bloating Severity Item

The bloating severity item is a standalone item from the original GCSI-DD. This item refers to bloating as "feeling like you need to loosen your clothes" and asks patients to rate the severity of their bloating over the past 24 hours on a five-point Likert scale (None, Mild, Moderate, Severe, Very Severe). This item was included within the daily diary for EVA-20216-01, as is intended for daily administration in the study.

** This measure was used in EVA-20216-01.*

4.1.4 ANMS GCSI-DD Total Score

The ANMS GCSI-DD total score is generated by summing the scores on each of the five symptom items (nausea, vomiting, early satiety, postprandial fullness, and upper abdominal pain) and the bloating severity item and then dividing by 6 (i.e., the number of items within the core symptom composite score plus the bloating severity item).

** This measure was used in EVA-20216-01.*

4.1.5 Patient-reported Nausea Subscore

The nausea item from the ANMS GCSI-DD was used as a standalone item for the TAK-906-1002 trial inclusion criteria. This item refers to nausea as "feeling sick to your stomach as if you were going to vomit or throw up" and asks patients to rate the severity of their nausea over the past 24 hours on a five-point Likert scale (None, Mild, Moderate, Severe, Very Severe). Subjects needed to have a nausea subscale score of ≥2 (Moderate) on at least three of the seven days during Screening.

** This measure was used in TAK-906-1002.*

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

The Patient Assessment of Upper Gastrointestinal Disorders-Symptom Severity Index (PAGI-SYM) was developed to assess gastrointestinal symptoms in patients with gastroesophageal reflux disease (GERD), dyspepsia, and gastroparesis. The instrument includes 20 symptom questions in six subscales: heartburn/regurgitation, fullness/early satiety, nausea/vomiting, bloating, upper abdominal pain, and lower abdominal pain (Rentz et al 2004). The recall period is the past two weeks.

** This measure was used in EVA-20216-01 and TAK-906-1002.*

4.3 Patient Global Impression of Severity (PGI-S) and Patient Global Impression of Improvement (PGI-I)

The Patient Global Impression of Severity (PGI-S) is a one-item scale designed to assess a patient's impression of disease severity. This questionnaire includes a four-point Likert scale and is based on a patient's current symptoms. The response options include: Normal, Not at all ill; Mildly ill; Moderately ill; and Severely ill. Like the PGI-S, the Patient Global Impression of Improvement (PGI-I) is also a one-item scale, and is based on a seven-point Likert scale. The PGI-I was designed to assess if a patient's symptoms improved. The response options include: Very much improved, Much improved, A little improved, No change, A little worse, Much worse, and Very much worse.

** These measures were used in EVA-20216-01.*

4.4 Patient-Reported Outcomes Measurement Information System (PROMIS) Global Scale

The Patient-Reported Outcomes Measurement Information System (PROMIS) Global Scale (Hays et al. 2009) is a self-administered generic health status scale that includes 10 items representing physical health, pain, fatigue, mental health, social health, and overall health. Items are assessed on either a five-point response scale, ranging from 1 (Poor/Not at all/None/Never) to 5 (Excellent/Completely/Very severe/Always)—or an 11-point Likert type scale, where the response to the question ranges from 0 (No pain) to 10 (Worst pain imaginable).

** This measure was used in EVA-20216-01.*

4.5 Clinical Symptom Severity Rating (CSSR)

A CSSR form was completed by the clinical site study staff at the Baseline visit to capture information on overall severity of gastroparesis symptoms and severity ratings on selected symptoms (i.e., nausea, abdominal distension, bloating, early satiety, vomiting, and abdominal pain). Severity was rated on a 0 (None) to 4 (Very severe) scale (Revicki et al. 2003).

** This measure was used in EVA-20216-01 and TAK-906-1002.*

4.6 Clinician Global Impression-Severity (CGI-S) and Clinician Global Impression-Improvement (CGI-I)

The Clinician Global Impression-Severity (CGI-S) is based on observed and reported symptoms, behavior, and function on a seven-point scale. The CGI-S has a seven-day recall period, and the score reflects the average severity level across the seven days (Busner & Targum 2007). The Clinician Global Impression-Improvement (CGI-I) is also observer-reported and uses a seven-point scale. Unlike the CGI-S, the CGI-I compares the patient's overall condition each time the patient is seen after Baseline.

** These measures were used in EVA-20216-01.*

4.7 Gastric Emptying Breath Test (GEBT)

The Gastric Emptying Breath Test (GEBT) is a non-radioactive, non-invasive, orally administered test for measuring the rate of solid phase gastric emptying (from stomach to small intestine) in adults. The GEBT is conducted over a four-hour period after an overnight fast. Patients have an initial breath test and then eat a special test meal that includes Spirulina (a nutritional supplement), powdered egg, and Saltine crackers. The Spirulina contains a blue-green algae enriched with a non-radioactive material that can be measured in breath tests. The GEBT has been validated against the gold standard reference method of gastric scintigraphy.

** This measure was used in TAK-906-1002.*

4.8 SmartPill

SmartPill is an ingestible capsule that measures pressure, pH, and temperature as it travels through the gastrointestinal tract to assess gastrointestinal motility. It can measure combined small and large bowel transit time. The SmartPill wirelessly transmits data about the gastrointestinal tract to a recorder that is worn on a belt clip or lanyard around the neck as one goes about his/her daily activities. The SmartPill eliminates radiation exposure and is the only motility test that provides a complete transit profile of the gastrointestinal tract.

** This measure was used in TAK-906-1002.*

Table C. Relevant Measures in EVA-20216-01 and TAK-906-1002

Measure	EVA-20216-01	TAK-906-1002
Medical history and demographics	Baseline	Screening
Study drug administration	No	Yes
GEBT		Day -1, Day 1, Day 7
SmartPill ingestion		Day -1 and Day 7
Nausea item score of ANMS GCSI-DD		7 days during Screening
ANMS CGSI-DD	Daily from Baseline through Day 28	Daily from Screening Day -7 through Day -1, and from Day 1 through Day 9 (or Early Termination)
Patient-rated bloating severity item	Daily from Baseline through Day 28	
PAGI-SYM	Baseline and Day 28	Pre-dose Day 1 and Day 9
PGI-S	Baseline and Day 28	
PGI-I	Day 28	
PROMIS Global Scale	Baseline and Day 28	
CSSR	Baseline and Day 28	Pre-dose Day 1 and Day 9
CGI-S	Baseline and Day 28	
CGI-I	Day 28	

GEBT = gastric emptying breath test; ANMS GCSI-DD = American Neurogastroenterology and Motility Society Gastroparesis Cardinal Symptom Index-Daily Diary; PAGI-SYM = Patient Assessment of Upper Gastrointestinal Disorders-Symptom Severity Index; PGI-S = Patient Global Impression of Severity; PGI-I = Patient Global Impression of Improvement; PROMIS = Patient-Reported Outcomes Measurement Information System; CSSR = Clinical Symptom Severity Rating; CGI-S = Clinician Global Impression-Severity; CGI-I = Clinician Global Impression-Improvement

5.0 Statistical Methods

5.1 General Statistical Procedures

All analyses will follow this approved SAP. All work will be subject to quality control and documentation procedures to ensure that the analyses and final results are accurate and thorough, and that the analyses can be reproduced. As the data are analyzed, some deviation from planned analysis may be anticipated (e.g., missing data, small sample size). In the case of substantial deviations from data expectations that would make the proposed analysis inappropriate (e.g., sample size or missing data issues), Evidera and Takeda will work together to determine the necessary revisions in the analyses. Deviations, revisions, and the rationale for the change(s) will be summarized.

Of note, EVA-20216-01 is enrolling patients aged 18-85, with a BMI ≥18 and ≤35kg/m^2, while TAK-906-1002 is enrolling patients aged 18-75, with a BMI ≥18 and ≤40kg/m2. Should the sample size of patients aged 75-85 in EVA-20216-01 exceed N=10, the ANMS-GCSI-DD will be examined descriptively and

compared to the rest of the population for consistency to determine is psychometric testing should be completed in this subgroup.

All statistical tests will use a two-sided significance level of 0.05 unless otherwise noted.

5.1.1 Missing Data – ANMS-GCSI-DD

For the ANMS GCSI-DD scoring, the following data rules will be applied:

- If more than 20% of the items are missing, the daily endpoint, composite, and total scores will be set to missing
- If less than four days (<4) of diary data are available for a given week, that week's scores will be considered missing

5.2 Descriptive Statistics

Baseline sociodemographic and clinical characteristics will be presented using descriptive statistics (mean, median, standard deviation [SD], range) (Table 1).

The weekly ANMS GCSI-DD endpoint, composite, and total scores, and individual item scores will be summarized to examine floor and ceiling effects (Table 2). This will be completed using Day 1 to Day 7 data from EVA-20216-01 and from Screening Day -7 to Day -1 data from TAK-906-1002. Any item that shows a floor (>25% have the minimum response) or ceiling effect (>25% have the maximum response) will be flagged.

Change scores for the ANMS GCSI-DD endpoint, composite, total score, and individual item scores will also be summarized (Table 3). The change score will be calculated by subtracting the mean Time 1 score from the mean Time 2 score. Time 1 and Time 2 will be defined in the following ways for the two datasets:

- EVA-20216-01
  - Time 1: Day 1 to Day 7
  - Time 2: Day 22 to Day 28
- TAK-906-1002
  - Time 1: Screening Day -7 to Day -1
  - Time 2: Day 2 to Day 8

5.3 Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) will be conducted to assess the unidimensionality of the ANMS GCSI-DD endpoint, composite, and total scores. This analysis will be conducted using a random day selected during Day 1 to Day 7 from EVA-20216-01 and from Screening Day -7 to Day -1 from TAK-906-1002 (Table 4).

Several fit statistics will be used to provide information about the adequacy of the model to explain the data. Model fit will be assessed with the comparative fit index (CFI) and root mean square error of

approximation (RMSEA). In general, the model will be considered to have good fit and explain the data well if the CFI is ≥0.90 (Hu & Bentler 1999). RMSEA is a measure of fit assessing the discrepancy between the predicted and observed data per degree of freedom; values <0.07 are generally considered acceptable (Browne & Cudeck 1993), and the 90% CI for RMSEA should be narrow—thereby giving additional confidence in the estimate. For simple models, the RMSEA can be inflated, thereby giving the impression of poor fit (Breivik & Olsson 2001). Adequacy of item fit will also be assessed through the examination of modification indices, item residual correlations, and item factor loadings. For ill-fitting models, the team will identify possible reasons for the misfit and adapt the models, if necessary.

5.3.1 Item Response Theory

Following the CFA analyses, item response theory (IRT) analysis will be used to examine whether each item in the ANMS GCSI-DD endpoint, composite, and total score exhibits psychometric properties with the following criteria: 1) item response options well ordered; and 2) items form a unidimensional construct. The IRT analyses will be completed for the combined EVA-20216-01 and TAK-906-1002 study populations.

The Samejima's graded response model (Samejima 1969), which allows for analyses of measures with Likert-type scales, will be used for the IRT analyses. With this model, for each item, the item parameters are the slope parameter, ai, and the m-1 category threshold parameters. These parameters describe the items in relationship with the underlying latent construct; i.e., gastroparesis symptom severity. The S-X2 fit statistics (Orlando & Thissen 2000) will be computed for each item to assess the model fit, and the marginal reliability will be computed to assess internal reliability (analogous to Cronbach's alpha). The value of p<0.001 will be used to indicate mis-fitting items (Table 5).

5.4 Reliability

5.4.1 Internal Consistency Reliability

Internal consistency reliability addresses the extent to which individual items within the ANMS GCSI-DD endpoint, composite, and total score are related to each other; item-total correlations will also be examined. Internal consistency reliability will be assessed by Cronbach's coefficient using Baseline data. There are no tests of statistical significance; the values are presented descriptively on an interval-level scale from 0–1.0, with higher scores indicating a more reliable (homogeneous) instrument. Coefficients of >0.7 will indicate good internal consistency reliability (Nunnally & Bernstein 1994; Cronbach 1951). This analysis will be conducted using a random day selected during Day 1 to Day 7 from EVA-20216-01 and from Screening Day -7 to Day -1 from TAK-906-1002 (Table 6).

5.4.2 Test-retest Reliability

Test-retest reliability will be examined to assess the stability of the ANMS GCSI-DD endpoint, composite, total score, and individual items over time within a stable population (Table 7 series). The following plans for defining a stable patient population are proposed:

- Patients that change 0 units from Time 1 to Time 2 on the physician overall rating of gastroparesis severity from the CSSR will be defined as stable. For EVA-20216-01, Time 1 is Day 1, and Time 2 is

Day 28. For TAK-906-1002, Time 1 is Pre-dose Day 1, and Time 2 is Day 9. If the sample size is insufficient, then patients that change 0 or 1 unit from Time 1 to Time 2 will be considered stable.

- Patients that change 0 or 1 units from Time 1 to Time 2 on the self-rated overall gastroparesis symptoms severity from the ANMS GCSI-DD will be defined as stable. For EVA-20216-01, Time 1 is Day 1, and Time 2 is Day 28. For TAK-906-1002, Time 1 is Day -1, and Time 2 is Day 8.
- Patients that rate their change in symptoms as "no change" from Time 1 to Time 2 on the PGI-S will be defined as stable. This analysis will be completed with EVA-20216-01 data only; Time 1 is Day 1, and Time 2 is Day 28.
- Patients with clinician ratings of change in symptoms as "no change" from Time 1 to Time 2 on the CGI-S will be defined as stable. This analysis will be completed with EVA-20216-01 data only; Time 1 is Day 1, and Time 2 is Day 28.

The assessment of test-retest reliability of the ANMS GCSI-DD endpoint, composite, total score, and individual item scores will be conducted by comparing the stable subjects' scores at Time 1 and Time 2 using paired t-tests and intra-class correlations coefficients (ICCs).

- EVA-20216-01
  - Time 1: Day 1 to Day 7
  - Time 2: Day 22 to Day 28
- TAK-906-1002
  - Time 1: Screening Day -7 to Day -1
  - Time 2: Day 2 to Day 8

ICCs range from 0–1.0, with higher scores indicating a more stable instrument. The hypothesis is that there will be no significant differences in scores when there is no change in disease status. Landis and Koch (1977) characterized ICC values <0 as indicating no agreement, 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.0 as almost perfect agreement.

5.5 Validity

Validity refers to the extent to which the instrument measures what it is intended to measure, and is frequently tested by examining correlations with other indicators of similar/related constructs. Validity of the ANMS GCSI-DD endpoint, composite, total scores, and individual items will be assessed for convergent validity and known-groups validity at Baseline.

5.5.1 Convergent Validity

Convergent validity involves demonstrating that different measures of the same concept substantially correlate. Using Pearson's product-moment and Spearman's rank correlation coefficients, the relationship between the ANMS GCSI-DD endpoint, composite, total scores, and individual items with the following variables will be examined (Tables 8.1 and 8.2):

- EVA-20216-01 (Baseline and Day $28^{+/-}1$)
  - Patient-rated bloating severity
  - PAGI-SYM Scores
  - PGI-S
  - PROMIS Global Scale Scores
  - CSSR
  - CGI-S
- TAK-906-1002 (Day -1/Pre-dose Day 1 and Day $8^{+/-}1$)
  - GEBT
  - SmartPill
  - PAGI-SYM Scores
  - CSSR

A correlation coefficient >0.3 indicates moderate convergent validity, whereas a correlation coefficient >0.5 indicates strong convergent validity (Cohen 1988).

The ANMS GCSI-DD endpoint, composite, total scores, and individual item scores, and the PAGI-SYM and PGI-S are expected to be moderately to strongly correlated with each other.

5.5.2 Known-groups Validity

Known-groups validity is the extent to which scores from an instrument are distinguishable between groups of subjects that differ by a relevant clinical (or other) indicator. Determining known-groups validity involves evaluating an instrument in relation to clinical measures of disease status (Stewart et al. 1992). To evaluate known-groups validity, the ANMS GCSI-DD endpoint, composite, total scores, and individual items will be analyzed by the physician's overall rating of gastroparesis severity from the CSSR response option categories (None, Mild, Moderate, Severe, Very severe), the self-rated overall gastroparesis symptom severity from the ANMS GCSI-DD response option categories (None, Mild, Moderate, Severe, Very severe), and the PGI-S response option categories (Normal/Not at all ill, Mildly ill, Moderately ill, Severely ill) (EVA-20216-01 dataset only) (Tables 9.1, 9.2, and 9.3). For the EVA-20216-01 dataset, the analyses will be at Baseline and Day $28^{+/-}1$; for the TAK-906-1002 dataset, the analyses will be at Day -1/Pre-dose Day 1 and Day $8^{+/-}1$. The analysis of covariance (ANCOVA) model will include the ANMS GCSI-DD endpoint, composite, total score, and individual items as the dependent variable, and the known-group criterion variable as the independent variable.

5.6 Exploratory Analyses

Data from the EVA-20216-01 dataset and TAK-906-1002 study population may be used to evaluate responsiveness and responder definition thresholds of the ANMS GCSI-DD endpoint, composite, total scores, and individual items. Execution of these analyses will be dependent on the number of patients enrolled in EVA-20216-01 that are initiating new therapy, adequate sample sizes in EVA-20216-01 within

change groups from the Day 1 to Day 28 window, and adequate sample sizes in TAK-906-1002 within change groups from the Day -1 to Day 9 window. The proposed analyses are described below.

5.6.1 Responsiveness

Responsiveness refers to the extent to which the instrument can detect true change in participants known to have changed in clinical status (Hays & Revicki 2005). To explore the responsiveness of the ANMS GCSI-DD endpoint, composite, total score, and individual items, the association between change scores and changes in the CSSR and PGI-S (EVA-20216-01 only) and PGI-I at Day $28^{+/-}1$ (EVA-20216-01 only) will be evaluated (Tables 10, 11, and 12). Improvement for each of the measures will be determined based on the distribution of the changes in the variables and clinical relevance. Regression models will be used to estimate the relationship between mean ANMS GCSI-DD endpoint, composite, total, and individual item change scores controlling for key covariates—including baseline score and demographic variables (e.g., age, gender, diabetic vs. idiopathic gastroparesis).

The responsiveness of the ANMS GCSI-DD endpoint, composite, total score, and individual items will be evaluated for changes from Time 1 to Time 2:

- EVA-20216-01
  - ANMS GCSI-DD Time 1: Day 1 to Day 7
  - ANMS GCSI-DD Time 2: Day 22 to Day 28
  - PGI-S and CSSR Time 1: Day 1
  - PGI-S, CSSR, and PGI-I Time 2: Day 28
- TAK-906-1002
  - ANMS GCSI-DD Time 1: Screening Day -7 to Day -1
  - ANMS GCSI-DD Time 2: Day 2 to Day 8
  - CSSR Time 1: Pre-Dose Day 1
  - CSSR Time 2: Day 9

5.6.2 Responder Definitions

The responder definition represents the individual patient PRO score change threshold over a pre-determined period of time that should be interpreted as a treatment benefit (FDA 2009). To explore the responder definition threshold estimates, anchor- and distribution-based estimates will be calculated. In addition, the cumulative distribution function (CDF) (a continuous plot of each individual subject's change scores) will be generated. Through the process of triangulation, the anchor- and distribution-based results and CDF will be examined, and a single responder definition threshold may be determined. As stated under Section 5.6, these analyses are dependent on adequate sample sizes.

5.6.2.1 Anchor-based Analysis

For the ANMS GCSI-DD endpoint, composite, total score, and individual items, patients will be categorized based on the change in CSSR, PGI-S, and PGI-I (EVA-20216-01 only) from Time 1 to Time 2. The mean

change in the ANMS GCSI-DD endpoint, composite, and total score, and individual items observed in the small improvement group ("-1") will be examined as a key anchor-based indicator of a responder (Table 13 and 14). The small improvement definitions for the CSSR, PGI-S, and PGI-I are below.

- CSSR: small improvement patients are those that improve by one category (i.e., from "moderate' to "mild")
- PSI-S: small improvement patients are those that improve by one category (i.e., from "severely ill" to "moderately ill")
- PGI-I: small improvement patients are those that endorse "a little improved"

5.6.2.2 Distribution-based Analysis

The first distribution-based approach will be a calculation of standard deviation units; the half standard deviation has been shown to provide a reasonable approximation of a meaningful change in PRO instruments (Norman et al. 2003). Half standard deviation and one-third standard deviation will be calculated for the ANMS GCSI-DD endpoint, composite, total score, and individual items. The standard error of measurement (SEM) (Wyrwich et al. 1999) establishes cutoffs for clinically significant change. The SEM is expressed in the instrument's original metric, which can facilitate ease of interpretation. The SEM will be computed as the standard deviation of an observed score related to its reliability (SD * sqrt of [1 - ICC]). The ICC used for the ANMS GCSI-DD endpoint, composite, total score, and individual items will be the mean of the estimates obtained in the reliability analyses. The SEM will be calculated for Time 1 and Time 2 (Table 15).

5.6.2.3 Triangulation

A single estimate of meaningful change scores of the ANMS GCSI-DD will be achieved by triangulating findings of the anchor- and distribution-based analyses. Triangulation has three iterative stages (Revicki et al. 2008; Leidy & Wyrwich 2005; Revicki et al. 2006):

1. Examining the range of anchor- and distribution-based estimates
2. Examining all the actual change scores for all individual participants at the relevant timepoints
3. Looking for gaps (spaces) in the observed change scores in the range identified in the anchor- and distribution-based estimates that create natural candidates for a cut-point that identifies important levels of disease progression or improvement

5.6.2.4 CDF

The CDF assists in the interpretation of the treatment effect across a range of possible disease responder definitions. The CDFs will be plotted separately for the ANMS GCSI-DD endpoint, composite, total score, and individual items. The CDF is a continuous plot of each individual subject's change plotted in order from greatest improvement to greatest worsening (from Time 1 to Time 2) on the x-axis, and the cumulative percent of patients experiencing that change on the y-axis.

The CDF plots will include an indicator line at the proposed responder definition that was triangulated based on anchor- and distribution-based analyses (Figures 1-10).

6.0 References

Breivik E, Olsson UH. Adding variables to improve fit: the effect of model size on fit assessment in LISREL. In: Cudeck R, du Toit S, Sörbom D, eds. *Structural Equation Modeling: Present and Future*. Lincolnwood, IL: Scientific Software International; 2001:169-194.

Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, eds. *Testing Structural Equation Models*. Beverly Hills, CA: Sage Company; 1993:136-162.

Busner J, Targum SD. The Clinical Global Impressions Scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)*. 2007;4(7):28-37.

Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

Cronbach LJ. Coefficient alpha and the internal structure of test. *Psychometrika*. 1951;16(3):297-334.

Food and Drug Administration. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. *Federal Register*. 2009;74(235):65132-65133.

Food and Drug Administration. Gastroparesis: Clinical Evaluation of Drugs for Treatment. Guidance for Industry. July 2015. Available at: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm455645.pdf. Access date: May 5, 2017.

Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Qual Life Res*. 2009;18(7):873-880.

Hays RD, Revicki DA. Reliability and validity, including responsiveness. In: Fayers PM, Hays RD, eds. *Assessing Quality of Life in Clinical Trials*. 2nd ed. New York: Oxford University Press; 2005:25-39.

Hu L, Bentler PM. Cutoff criterion for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling.* 1999;6:1-55.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.

Leidy NK, Wyrwich KW. Bridging the gap: using triangulation methodology to estimate minimal clinically important differences (MCIDs). *COPD*. 2005;2(1):157-165.

Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41(5):582-592.

Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd ed. New York, NY: McGraw-Hill; 1994.

Orlando M, Thissen D. Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*. 2000;24:50-64.

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. *Quality of Life Research*. 2004; 13(10):1737-1749.

Revicki DA, Rentz AM, Dubois D, Kahrilas P, Stanghellini V, Talley NJ, Tack J. Development and validation of a patient-assessed gastroparesis symptom severity measure: the Gastroparesis Cardinal Symptom Index. *Aliment Pharmacol Ther*. 2003;18(1):141-150.

Revicki DA, Cella D, Hays RD, et al. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes*. 2006;4:70.

Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61(2):102-109.

Samejima F. Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monographs*. 1969;34(4):Pt 2: whole No. 17.

Stewart AL, Hays RD, Ware JE. *Methods of Validating MOS Health Measures*. North Carolina: Duke University Press; 1992.

Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol*. 1999;52(9):861-873.

Tables

Table 1. Sociodemographic and Clinical Characteristics

	Total (N=)	EVA-20216-01 (N=)	TAK-906-1002 (N=)
Age			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Sex			
Female, n (%)			
Male, n (%)			
Missing, n (%)			
Race			
American Indian or Alaska Native, n (%)			
Asian, n (%)			
Black or African American, n (%)			
Native Hawaiian or Pacific Islander, n (%)			
White, n (%)			
Missing, n (%)			
Ethnicity			
Hispanic or Latino, n (%)			
Not Hispanic or Latino, n (%)			
Missing, n (%)			
Current Marital Status			
Single/never married, n (%)			
Married or living in marriage-like relationship, n (%)			
Widowed/separated/divorced/annulled, n (%)			
Missing, n (%)			
Current Living/Domestic Situation			
Living alone, n (%)			
Living with a partner or spouse, family, or friends, n (%)			
Other, n (%)			
Missing, n (%)			
Employment Status			

	Total (N=)	EVA-20216-01 (N=)	TAK-906-1002 (N=)
Employed full-time, n (%)			
Employed part-time, n (%)			
Homemaker, n (%)			
Student, n (%)			
Unemployed, n (%)			
Retired, n (%)			
Disabled, n (%)			
Missing, n (%)			
Highest Level of Education			
Elementary/primary school, n (%)			
Secondary/high school/GED, n (%)			
Some college or post-high school education training, n (%)			
College, n (%)			
Postgraduate degree, n (%)			
Missing, n (%)			
Years Since Gastroparesis Diagnosis			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Diabetes			
Type 1, n (%)			
Type 2, n (%)			
No, n (%)			
Current Medications and Treatments for Gastroparesis			
Prokinetic agent, n (%)			
Antiemetic agent, n (%)			
Gastric antisecretory agent, n (%)			
Pain Medications, n (%)			
Psychotropic agent, n (%)			
Diabetes treatment, n (%)			
Gastric electric stimulation (pacemaker) , n (%)			
Gastric surgery, n (%)			
Gastrostomy tube (G tube) , n (%)			

	Total (N=)	EVA-20216-01 (N=)	TAK-906-1002 (N=)
Jejunostomy tube (J tube) , n (%)			
Central line or PICC line, n (%)			
Other, n (%)			
Not on treatment, n (%)			
Prior Gastrointestinal Surgeries			
Fundoplication, n (%)			
Gastrectomy, n (%)			
Gastric banding, n (%)			
Gastric bypass, n (%)			
Gastric electric stimulation (GES), n (%)			
Jejunostomy (J-tube), n (%)			
Pyloroplasty, n (%)			
Vagotomy, n (%)			
Other, n (%)			
No prior surgeries, n (%)			
Comorbidities			
Auto-immune disorder, n (%)			
Cancer, n (%)			
Cardiovascular condition, n (%)			
Dyslipidemia/hyperlipidemia, n (%)			
Insomnia/sleep problems, n (%)			
Liver disease, n (%)			
Mood disorder or mental health condition, n (%)			
Neurological condition, n (%)			
Obesity, n (%)			
Respiratory condition, n (%)			
Other gastrointestinal disorder, n (%)			
None, n (%)			

Table 2. Item Level Characteristics: ANMS GCSI-DD

	Total (N=)	EVA-20216-01 (N=)	TAK-906-1002 (N=)
ANMS GCSI-DD Endpoint Score			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Floor, n (%)			
Ceiling, n (%)			
ANMS GCSI-DD Composite Score			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Floor, n (%)			
Ceiling, n (%)			
ANMS GCSI-DD Total Score			
Mean (SD)			-
Median (Q1-Q3)			-
Range (Min, Max)			-
Missing, n (%)			-
Floor, n (%)			-
Ceiling, n (%)			-
1. Nausea (feeling sick to your stomach as if you were going to vomit or throw up)			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Floor, n (%)			
Ceiling, n (%)			
2. Not able to finish a normal-sized meal (for a healthy person)			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			

	Total (N=)	EVA-20216-01 (N=)	TAK-906-1002 (N=)
Missing, n (%)			
Floor, n (%)			
Ceiling, n (%)			
3. Feeling excessively full after meals			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Floor, n (%)			
Ceiling, n (%)			
4. Upper abdominal pain (above the navel)			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Floor, n (%)			
Ceiling, n (%)			
5. Episodes of vomiting			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Floor, n (%)			
Ceiling, n (%)			
Bloating (feeling like you need to loosen your clothes)*			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Floor, n (%)			
Ceiling, n (%)			
Overall severity of your gastroparesis symptoms*			
Mean (SD)			
Median (Q1-Q3)			

	Total (N=)	EVA-20216-01 (N=)	TAK-906-1002 (N=)
Range (Min, Max)			
Missing, n (%)			
Floor, n (%)			
Ceiling, n (%)			

*Not included in the endpoint or composite score

Table 3. Item Change Scores from Time 1 to Time 2: ANMS GCSI-DD

	Total (N=)	EVA-20216-01 (N=)	TAK-906-1002 (N=)
Change in ANMS GCSI-DD Endpoint Score			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Change in ANMS GCSI-DD Composite Score			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Change in ANMS GCSI-DD Total Score			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Change in 1. Nausea (feeling sick to your stomach as if you were going to vomit or throw up)			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Change in 2. Not able to finish a normal-sized meal (for a healthy person)			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Change in 3. Feeling excessively full after meals			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Change in 4. Upper abdominal pain (above the navel)			
Mean (SD)			

	Total (N=)	EVA-20216-01 (N=)	TAK-906-1002 (N=)
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Change in 5. Episodes of vomiting			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Change in Bloating (feeling like you need to loosen your clothes)*			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			
Change in Overall severity of your gastroparesis symptoms*			
Mean (SD)			
Median (Q1-Q3)			
Range (Min, Max)			
Missing, n (%)			

*Not included in the endpoint or composite score

Table 4. Item Analysis: Confirmatory Factor Analysis for ANMS GCSI-DD

Item	Standardized Coefficient	Model Fit Statistics CFI = RMSEA (95% CI) = () WRMR =
ANMS GCSI-DD Endpoint Score		
1. Nausea (feeling sick to your stomach as if you were going to vomit or throw up)		
2. Not able to finish a normal-sized meal (for a healthy person)		
3. Feeling excessively full after meals		
4. Upper abdominal pain (above the navel)		
ANMS GCSI-DD Composite Score		
1. Nausea (feeling sick to your stomach as if you were going to vomit or throw up)		
2. Not able to finish a normal-sized meal (for a healthy person)		
3. Feeling excessively full after meals		
4. Upper abdominal pain (above the navel)		
5. Episodes of vomiting		
ANMS GCSI-DD Total Score		
1. Nausea (feeling sick to your stomach as if you were going to vomit or throw up)		
2. Not able to finish a normal-sized meal (for a healthy person)		
3. Feeling excessively full after meals		
4. Upper abdominal pain (above the navel)		
5. Episodes of vomiting		
6. Bloating (feeling like you need to loosen your clothes)		

Table 5. Item Response Theory Analysis: ANMS GCSI-DD Item Fit Statistics at Baseline

Items	Marginal reliability	Slope	Category Threshold Parameters					Model Fit		
			β_1 (SE)	β_2 (SE)	β_3 (SE)	β_4 (SE)	β_5 (SE)	DF	X2	Probability X2
ANMS GCSI-DD Endpoint Score										
1. Nausea (feeling sick to your stomach as if you were going to vomit or throw up)						-	-			
2. Not able to finish a normal-sized meal (for a healthy person)						-	-			
3. Feeling excessively full after meals						-	-			
4. Upper abdominal pain (above the navel)						-	-			
ANMS GCSI-DD Composite Score										
1. Nausea (feeling sick to your stomach as if you were going to vomit or throw up)							-			
2. Not able to finish a normal-sized meal (for a healthy person)							-			
3. Feeling excessively full after meals							-			
4. Upper abdominal pain (above the navel)							-			
5. Episodes of vomiting							-			
ANMS GCSI-DD Total Score										
1. Nausea (feeling sick to your stomach as if you were going to vomit or throw up)										
2. Not able to finish a normal-sized meal (for a healthy person)										
3. Feeling excessively full after meals										
4. Upper abdominal pain (above the navel)										
5. Episodes of vomiting										
6. Bloating (feeling like you need to loosen your clothes)										

Table 6. Internal Consistency Reliability: ANMS GCSI-DD

	Total (N=)		EVA-20216-01 (N=)		TAK-906-1002 (N=)	
	Deleted Item Correlation with Domain	Cronbach's α	Deleted Item Correlation with Domain	Cronbach's α	Deleted Item Correlation with Domain	Cronbach's α
ANMS GCSI-DD Endpoint Score						
1. Nausea (feeling sick to your stomach as if you were going to vomit or throw up)						
2. Not able to finish a normal-sized meal (for a healthy person)						
3. Feeling excessively full after meals						
4. Upper abdominal pain (above the navel)						
ANMS GCSI-DD Composite Score						
1. Nausea (feeling sick to your stomach as if you were going to vomit or throw up)						
2. Not able to finish a normal-sized meal (for a healthy person)						
3. Feeling excessively full after meals						
4. Upper abdominal pain (above the navel)						
5. Episodes of vomiting						
ANMS GCSI-DD Composite Score						
1. Nausea (feeling sick to your stomach as if you were going to vomit or throw up)					-	-
2. Not able to finish a normal-sized meal (for a healthy person)					-	-
3. Feeling excessively full after meals					-	-
4. Upper abdominal pain (above the navel)					-	-
5. Episodes of vomiting					-	-
6. Bloating (feeling like you need to loosen your clothes)					-	-

Table 7.1 Test-Retest Reliability of ANMS GCSI-DD Endpoint Score from Time 1 to Time 2

	N	Time 1 Mean (SD)	Time 2 Mean (SD)	Difference[1]	p-value[2]	ICC[3]
CSSR Stable Definition[4]						
EVA-20216-01 dataset						
TAK-906-1002						
Combined datasets						
ANMS GCSI-DD Stable Definition[5]						
EVA-20216-01 dataset						
TAK-906-1002						
Combined datasets						
PGI-S Stable Definition[6]						
EVA-20216-01 dataset						
CGI-S Stable Definition[7]						
EVA-20216-01 dataset						

[1] Mean difference= Time 2 - Time 1

[2] p-value for paired t-test

[3] Intraclass correlation coefficient

[4] Patients that change 0 from Time 1 to Time 2 on the physician overall rating of gastroparesis severity from the CSSR will be defined as stable.

[5] Patients that change 0 or 1 units from Time 1 to Time 2 on the self-rated overall gastroparesis symptoms severity from the ANMS GCSI-DD will be defined as stable.

[6] Patients that have “no change” from Time 1 to Time 2 on the PGI-S will be defined as stable. This analysis will be completed with EVA-20216-01 data only.

[7] Patients that have “no change” from Time 1 to Time 2 on the CGI-S will be defined as stable. This analysis will be completed with EVA-20216-01 data only.

Table 8.1. Correlations Between ANMS GCSI-DD and Other Measures at Baseline and Day $28^{+/-}1$ (EVA-20216-01)

	Correlations* Baseline/Day $28^{+/-}1$					
	Patient-rated bloating severity	PAGI-SYM	PGI-S	PROMIS Global Scale	CSSR	CGI-S
ANMS GCSI-DD Endpoint Score						
ANMS GCSI-DD Composite Score						
ANMS GCSI-DD Total Score						
Nausea						
Early Satiety						
Post-Prandial Fullness						
Upper Abdominal Pain						
Vomiting						
Bloating	-/-					
Overall Severity of Gastroparesis Symptoms						

* Spearman's correlation coefficients reported for continuous variables and Point bi-serial correlations reported for dichotomous variables: [1]p<0.0001, [2]p<0.001, [3]p<0.05

Table 8.2. Correlations Between ANMS GCSI-DD and Other Measures at Pre-dose Day 1 and Day $8^{+/-}1$ (TAK-906-1002)

	Correlations* Day 1/Day $8^{+/-}1$			
	GEBT	SmartPill	PAGI-SYM	CSSR
ANMS GCSI-DD Endpoint Score				
ANMS GCSI-DD Composite Score				
Nausea				
Early Satiety				
Post-Prandial Fullness				
Upper Abdominal Pain				
Vomiting				
Overall Severity of Gastroparesis Symptoms				

* Spearman's correlation coefficients reported for continuous variables and Point bi-serial correlations reported for dichotomous variables: [1]p<0.0001, [2]p<0.001, [3]p<0.05

Table 9.1. Known Groups Validity: ANMS GSCI-DD at Baseline and Day $28^{+/-}1$ (EVA-20216-01) and Pre-dose Day 1 and Day $8^{+/-}1$ (TAK-906-1002) using CSSR Global Severity Rating

ANMS GCSI-DD	CSSR Global Severity						
	None	Mild	Moderate	Severe	Very Severe	Overall F value (P value)[1]	Pairwise comparison[2]
ANMS GCSI-DD Endpoint Score							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
ANMS GCSI-DD Composite Score							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
ANMS GCSI-DD Total Score							
EVA-20216-01							
Nausea							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
Early Satiety							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
Post-Prandial Fullness							
EVA-20216-01							
TAK-906-1002							

ANMS GCSI-DD	CSSR Global Severity						
	None	Mild	Moderate	Severe	Very Severe	Overall F value (P value)[1]	Pairwise comparison[2]
Combined datasets							
Upper Abdominal Pain							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
Vomiting							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
Bloating							
EVA-20216-01							
Overall Severity of Gastroparesis Symptoms							
EVA-20216-01							
TAK-906-1002							
Combined datasets							

[1] An analysis of covariance (ANCOVA) model adjusting for age and gender

[2] Pairwise comparisons between LS means were performed using Bonferroni test adjusting for multiple comparisons

1: Group 1 vs. Group 2; 2: Group 1 vs. Group 3; 3: Group 1 vs. Group 4; 4: Group 1 vs. Group 5; 5: Group 2 vs. Group 3; 6: Group 2 vs. Group 4; 7: Group 2 vs. Group 5; 8: Group 3 vs. Group 4; 9: Group 3 vs. Group 5; 10: Group 4 vs. Group 5

Abbreviations: LS=Least squares; SE=Standard error

Table 9.2. Known Groups Validity: ANMS GSCI-DD at Baseline and Day $28^{+/-}1$ (EVA-20216-01) and Pre-dose Day 1 and Day $8^{+/-}1$ (TAK-906-1002) using ANMS GCSI-DD Global Severity Rating

ANMS GCSI-DD	ANMS GCSI-DD Global Severity						
	None	Mild	Moderate	Severe	Very Severe	Overall F value (P value)[1]	Pairwise comparison[2]
ANMS GCSI-DD Endpoint Score							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
ANMS GCSI-DD Composite Score							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
ANMS GCSI-DD Total Score							
EVA-20216-01							
Nausea							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
Early Satiety							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
Post-Prandial Fullness							
EVA-20216-01							
TAK-906-1002							

ANMS GCSI-DD	ANMS GCSI-DD Global Severity						
	None	Mild	Moderate	Severe	Very Severe	Overall F value (P value)[1]	Pairwise comparison[2]
Combined datasets							
Upper Abdominal Pain							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
Vomiting							
EVA-20216-01							
TAK-906-1002							
Combined datasets							
Bloating							
EVA-20216-01							

[1] An analysis of covariance (ANCOVA) model adjusting for age and gender

[2] Pairwise comparisons between LS means were performed using Bonferroni test adjusting for multiple comparisons

1: Group 1 vs. Group 2; 2: Group 1 vs. Group 3; 3: Group 1 vs. Group 4; 4: Group 1 vs. Group 5; 5: Group 2 vs. Group 3; 6: Group 2 vs. Group 4; 7: Group 2 vs. Group 5; 8: Group 3 vs. Group 4; 9: Group 3 vs. Group 5; 10: Group 4 vs. Group 5

Abbreviations: LS=Least squares; SE=Standard error

Table 9.3. Known Groups Validity: ANMS GSCI-DD at Baseline and Day $28^{+/-}1$ (EVA-20216-01) using PGI-S

ANMS GCSI-DD	PGI-S					
	Normal, not at all ill	Mildly ill	Moderately ill	Severely ill	Overall F value (P value)[1]	Pairwise comparison[2]
ANMS GCSI-DD Endpoint Score						
ANMS GCSI-DD Composite Score						
ANMS GCSI-DD Total Score						
Nausea						
Early Satiety						
Post-Prandial Fullness						
Upper Abdominal Pain						
Vomiting						
Bloating						
Overall Severity of Gastroparesis Symptoms						

[1] An analysis of covariance (ANCOVA) model adjusting for age and gender

[2] Pairwise comparisons between LS means were performed using Bonferroni test adjusting for multiple comparisons

1: Group 1 vs. Group 2; 2: Group 1 vs. Group 3; 3: Group 1 vs. Group 4; 4: Group 2 vs. Group 3; 5: Group 2 vs. Group 4; 6: Group 3 vs. Group 4

Abbreviations: LS=Least squares; SE=Standard error

Table 10. Responsiveness of the ANMS GCSI-DD by Change in CSSR

	Change in CSSR			p-value[1]
	Worsened (Δ+1 or greater) (N = XXX)	No change (N = XXX)	Improved (Δ-1 or lower) (N = XXX)	
ANMS GCSI-DD Endpoint Score	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
ANMS GCSI-DD Composite Score	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
ANMS GCSI-DD Total Score	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Nausea	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Early Satiety	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Post-Prandial Fullness	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Upper Abdominal Pain	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Vomiting	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Bloating	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Overall Severity of Gastroparesis Symptoms	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX

[1] General linear model (PROC GLM) controlling for age, gender, and DG.

Table 11. Responsiveness of the ANMS GCSI-DD by Change in PGI-S

	Change in PGI-S			p-value[1]
	Worsened (Δ+1 or greater) (N = XXX)	No change (N = XXX)	Improved (Δ-1 or lower) (N = XXX)	
ANMS GCSI-DD Endpoint Score	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
ANMS GCSI-DD Composite Score	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
ANMS GCSI-DD Total Score	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Nausea	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Early Satiety	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Post-Prandial Fullness	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Upper Abdominal Pain	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Vomiting	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Bloating	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Overall Severity of Gastroparesis Symptoms	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX

[1] General linear model (PROC GLM) controlling for age, gender, and DG.

Table 12. Responsiveness of the ANMS GCSI-DD by Day 28$^{+/-}$1 PGI-I

	Change in PGI-I			p-value[1]
	Worsened (Δ-1 or greater) (N = XXX)	No change (N = XXX)	Improved (Δ+1 or lower) (N = XXX)	
ANMS GCSI-DD Endpoint Score	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
ANMS GCSI-DD Composite Score	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
ANMS GCSI-DD Total Score	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Nausea	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Early Satiety	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Post-Prandial Fullness	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Upper Abdominal Pain	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Vomiting	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Bloating	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX
Overall Severity of Gastroparesis Symptoms	XX.X (XX.X)	XX.X (XX.X)	XX.X (XX.X)	0.XXX

[1] General linear model (PROC GLM) controlling for age, gender, and DG.

Table 13. Responder Definition: Change in ANMS GCSI-DD by Change in CSSR from Time 1 to Time 2

Change in CSSR		Change in ANMS GCSI-DD Endpoint Score	
		N	Mean (SD)
Worsening	+4	XXX	XX.X (X.X)
	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
	-4	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in ANMS GCSI-DD Composite Score	
Worsening	+4	XXX	XX.X (X.X)
	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
	-4	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in ANMS GCSI-DD Total Score	
Worsening	+4	XXX	XX.X (X.X)
	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
	-4	XXX	XX.X (X.X)
		P-value	<0.XXX

Change in CSSR		Change in ANMS GCSI-DD Endpoint Score	
		N	Mean (SD)
		Change in Nausea	
Worsening	+4	XXX	XX.X (X.X)
	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
	-4	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Early Satiety	
Worsening	+4	XXX	XX.X (X.X)
	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
	-4	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Post-Prandial Fullness	
Worsening	+4	XXX	XX.X (X.X)
	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
	-4	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Upper Abdominal Pain	

Change in CSSR		Change in ANMS GCSI-DD Endpoint Score	
		N	Mean (SD)
Worsening	+4	XXX	XX.X (X.X)
	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
	-4	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Vomiting	
Worsening	+4	XXX	XX.X (X.X)
	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
	-4	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Bloating	
Worsening	+4	XXX	XX.X (X.X)
	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
	-4	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Overall Severity of Gastroparesis Symptoms	
Worsening	+4	XXX	XX.X (X.X)

Change in CSSR		Change in ANMS GCSI-DD Endpoint Score	
		N	Mean (SD)
	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
	-4	XXX	XX.X (X.X)
		P-value	<0.XXX

Table 14. Responder Definition: Change in ANMS GCSI-DD by Change in PGI-S from Time 1 to Time 2

Change in PGI-S		Change in ANMS GCSI-DD Endpoint Score	
		N	Mean (SD)
Worsening	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in ANMS GCSI-DD Composite Score	
Worsening	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in ANMS GCSI-DD Total Score	
Worsening	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Nausea	
Worsening	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)

Change in PGI-S		Change in ANMS GCSI-DD Endpoint Score	
		N	Mean (SD)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Early Satiety	
Worsening	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Post-Prandial Fullness	
Worsening	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Upper Abdominal Pain	
Worsening	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Vomiting	
Worsening	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)

Change in PGI-S		Change in ANMS GCSI-DD Endpoint Score	
		N	Mean (SD)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Bloating	
Worsening	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Overall Severity of Gastroparesis Symptoms	
Worsening	+3	XXX	XX.X (X.X)
	+2	XXX	XX.X (X.X)
	+1	XXX	XX.X (X.X)
No Change	0	XXX	XX.X (X.X)
Improvement	-1	XXX	XX.X (X.X)
	-2	XXX	XX.X (X.X)
	-3	XXX	XX.X (X.X)
		P-value	<0.XXX

Table 15. Responder Definition: Change in ANMS GCSI-DD by PGI-I at Time 2

PGI-I		Change in ANMS GCSI-DD Endpoint Score	
		N	Mean (SD)
Worsening	Very much worse	XXX	XX.X (X.X)
	Much worse	XXX	XX.X (X.X)
	A little worse	XXX	XX.X (X.X)
No Change	No change	XXX	XX.X (X.X)
Improvement	A little improved	XXX	XX.X (X.X)
	Much improved	XXX	XX.X (X.X)
	Very much improved	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in ANMS GCSI-DD Composite Score	
Worsening	Very much worse	XXX	XX.X (X.X)
	Much worse	XXX	XX.X (X.X)
	A little worse	XXX	XX.X (X.X)
No Change	No change	XXX	XX.X (X.X)
Improvement	A little improved	XXX	XX.X (X.X)
	Much improved	XXX	XX.X (X.X)
	Very much improved	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in ANMS GCSI-DD Total Score	
Worsening	Very much worse	XXX	XX.X (X.X)
	Much worse	XXX	XX.X (X.X)
	A little worse	XXX	XX.X (X.X)
No Change	No change	XXX	XX.X (X.X)
Improvement	A little improved	XXX	XX.X (X.X)
	Much improved	XXX	XX.X (X.X)
	Very much improved	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Nausea	
Worsening	Very much worse	XXX	XX.X (X.X)
	Much worse	XXX	XX.X (X.X)
	A little worse	XXX	XX.X (X.X)
No Change	No change	XXX	XX.X (X.X)
Improvement	A little improved	XXX	XX.X (X.X)
	Much improved	XXX	XX.X (X.X)

PGI-I		Change in ANMS GCSI-DD Endpoint Score	
		N	Mean (SD)
	Very much improved	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Early Satiety	
Worsening	Very much worse	XXX	XX.X (X.X)
	Much worse	XXX	XX.X (X.X)
	A little worse	XXX	XX.X (X.X)
No Change	No change	XXX	XX.X (X.X)
Improvement	A little improved	XXX	XX.X (X.X)
	Much improved	XXX	XX.X (X.X)
	Very much improved	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Post-Prandial Fullness	
Worsening	Very much worse	XXX	XX.X (X.X)
	Much worse	XXX	XX.X (X.X)
	A little worse	XXX	XX.X (X.X)
No Change	No change	XXX	XX.X (X.X)
Improvement	A little improved	XXX	XX.X (X.X)
	Much improved	XXX	XX.X (X.X)
	Very much improved	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Upper Abdominal Pain	
Worsening	Very much worse	XXX	XX.X (X.X)
	Much worse	XXX	XX.X (X.X)
	A little worse	XXX	XX.X (X.X)
No Change	No change	XXX	XX.X (X.X)
Improvement	A little improved	XXX	XX.X (X.X)
	Much improved	XXX	XX.X (X.X)
	Very much improved	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Vomoting	
Worsening	Very much worse	XXX	XX.X (X.X)
	Much worse	XXX	XX.X (X.X)
	A little worse	XXX	XX.X (X.X)
No Change	No change	XXX	XX.X (X.X)

PGI-I		Change in ANMS GCSI-DD Endpoint Score	
		N	Mean (SD)
Improvement	A little improved	XXX	XX.X (X.X)
	Much improved	XXX	XX.X (X.X)
	Very much improved	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Bloating	
Worsening	Very much worse	XXX	XX.X (X.X)
	Much worse	XXX	XX.X (X.X)
	A little worse	XXX	XX.X (X.X)
No Change	No change	XXX	XX.X (X.X)
Improvement	A little improved	XXX	XX.X (X.X)
	Much improved	XXX	XX.X (X.X)
	Very much improved	XXX	XX.X (X.X)
		P-value	<0.XXX
		Change in Overall Severity of Gastroparesis Symptoms	
Worsening	Very much worse	XXX	XX.X (X.X)
	Much worse	XXX	XX.X (X.X)
	A little worse	XXX	XX.X (X.X)
No Change	No change	XXX	XX.X (X.X)
Improvement	A little improved	XXX	XX.X (X.X)
	Much improved	XXX	XX.X (X.X)
	Very much improved	XXX	XX.X (X.X)
		P-value	<0.XXX

Table 16. Responder Definitions: Distribution-Based Estimates

	Distribution-Based Responder Definitions		
	½ Baseline SD	1/3 Baseline SD	SEM[1]
ANMS GCSI-DD Endpoint Score			
Time 1	X.XX	X.XX	---
Time 1 to Time 2	---	---	X.XX
ANMS GCSI-DD Composite Score			
Time 1	X.XX	X.XX	---
Time 1 to Time 2	---	---	X.XX
ANMS GCSI-DD Total Score			
Time 1	X.XX	X.XX	---
Time 1 to Time 2	---	---	X.XX
Nausea			
Time 1	X.XX	X.XX	---
Time 1 to Time 2	---	---	X.XX
Early Satiety			
Time 1	X.XX	X.XX	---
Time 1 to Time 2	---	---	X.XX
Post-Prandial Fullness			
Time 1	X.XX	X.XX	---
Time 1 to Time 2	---	---	X.XX
Upper Abdominal Pain			
Time 1	X.XX	X.XX	---
Time 1 to Time 2	---	---	X.XX
Vomiting			
Time 1	X.XX	X.XX	---
Time 1 to Time 2	---	---	X.XX
Bloating			
Time 1	X.XX	X.XX	---
Time 1 to Time 2	---	---	X.XX
Overall Severity of Gastroparesis Symptoms			
Time 1	X.XX	X.XX	---
Time 1 to Time 2	---	---	X.XX

[1] SEM=baseline standard deviation of the measure multiplied by the square root of one minus its reliability coefficient.