Statistical Analysis Plan

NCT Number: NCT05353985

Title: A Phase 2, Randomized, Double-blind, Placebo-Controlled, Dose-Ranging Study to Evaluate the Efficacy and Safety of TAK-062 for the Treatment of Active Celiac Disease in Subjects Attempting a Gluten-Free Diet

Study Number: TAK-062-2001

Document Version and Date: Version: 4.0, 03 December 2024

Certain information within this document has been redacted (ie, specific content is masked irreversibly from view) to protect either personally identifiable information or company confidential information.

STATISTICAL ANALYSIS PLAN

Study Number: *TAK-062-2001*

Study Title: *A Phase 2, Randomized, Double-Blind, Placebo-Controlled, Dose-Ranging Study to Evaluate the Efficacy and Safety of TAK-062 for the Treatment of Active Celiac Disease in Subjects Attempting a Gluten-Free Diet*

Phase: *2*

Version: 4.0

Date: *03 December 2024*

Prepared by:

[redacted], MS.

[redacted], Statistics, Statistics and Quantitative Sciences (SQS)

Takeda Pharmaceuticals

Based on:

Protocol Version: Amendment No. 5

Protocol Date: -*16 Nov 2023*

REVISION HISTORY

Version	Approval Date	Primary Rationale for Revision
Original version 1	2 MAY 2022	Not Applicable
Version 2.0	27 FEB 2024	Updated based on protocol amendment 5
Version 3.0	23 AUG 2024	[redacted] Added additional exploratory endpoints.
Version 4.0	03 DEC 2024	[redacted]

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

ABBREVIATIONS

AE	adverse event
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
AUC	area under the curve
BMI	body mass index
CD-QOL	Celiac disease quality of life
CDAQ	Coeliac Disease Assessment Questionnaire
CDSD	Celiac Disease Symptom Diary
CeD	celiac disease
CI	confidence interval
COVID-19	coronavirus disease 2019
CPK	creatine phosphokinase
CSR	clinical study report
DGP	deaminated gliadin peptide
ECG	electrocardiogram
E-DMC	external data monitoring committee
EGD	esophagogastroduodenoscopy
eCRF	electronic case report form
EQ-5D-5L	EuroQol 5-Dimensions-5 Levels
FAS	full analysis set
FSH	follicle-stimulating hormone
GFD	gluten-free diet
GGT	γ-glutamyl transferase
GI	Gastrointestinal
GIP	gluten immunogenic peptide
GSRS	Gastrointestinal Symptom Rating Scale
HbA1c	glycosylated hemoglobin
HLA	human leukocyte antigen
HRQOL	health-related quality of life
IA	interim analysis
ICDSQ	Impact of Celiac Disease Symptoms Questionnaire
ICE	intercurrent events
ICF	informed consent form
ICH	International Conference on Harmonisation
ID	identification
IDMC	independent data monitoring committee
IEC	independent ethics committee
IEL	intraepithelial lymphocyte
IgA	Immunoglobulin A

IgG	Immunoglobulin G
INR	international normalized ratio
IRB	institutional review board
LLN	lower limit of normal
LS	least squares
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MMRM	mixed-effects model for repeated measures
mpVCIEL	Multi-parametric composite score of Vh:Cd and IEL
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PK	pharmacokinetic
PPS	per-protocol set
PRO	patient-reported outcomes
PROMIS	Patient-reported Outcomes Measurement Information System
PT	Preferred Term (MedDRA)
SAE	serious adverse event
SAF	safety analysis set
SAP	statistical analysis plan
SD	standard deviation
SF-12	Short Form 12-Item
SIGE	simulated inadvertent gluten exposure
SOC	System Organ Class
TEAE	treatment-emergent adverse event
[redacted]	[redacted]
[redacted]	[redacted]
tTG	tissue transglutaminase
ULN	upper limit of normal
[redacted]	[redacted]
VCIEL	Composite score of Vh:Cd ratio and IEL
Vh:Cd	villus height to crypt depth ratio
WHODrug	World Health Organization Drug Dictionary
WPAI+CIQ: CeD	Work Productivity and Activity Impairment plus Classroom Impairment Questionnaire: Celiac Disease

1.0 OBJECTIVES, ENDPOINTS AND ESTIMANDS

1.1 Objectives

1.1.1 Primary Objective

The primary objective of this study is:

- *To evaluate the efficacy of TAK-062, as measured by the CDSD, for reducing celiac-related symptoms due to gluten exposure in subjects with celiac disease (CeD) attempting to maintain a gluten-free diet (GFD) in treated subjects versus placebo controls.*

1.1.2 Secondary Objectives

Secondary objectives of this study are:

- *To evaluate the efficacy of TAK-062, as measured by Vh:Cd, for improvement of small intestine mucosal injury due to gluten exposure in subjects with CeD attempting to maintain a GFD in treated subjects versus placebo controls.*
- *To evaluate the safety and tolerability of TAK-062.*

1.1.3 Exploratory/Additional Objectives

Exploratory/additional objectives of this study are:

- *To further evaluate the psychometric properties of the Celiac Disease Symptom Diary (CDSD) in symptomatic subjects with CeD.*
- *To evaluate the efficacy of TAK-062 on reducing subjects most bothersome symptoms.*
- *To evaluate the treatment effect of TAK-062 on intraepithelial lymphocyte (IEL) counts and quantitative and qualitative histological measures of disease severity.*
- *To evaluate immunogenicity to TAK-062.*
- *To evaluate the effect of TAK-062 on quality of life and additional patient-reported outcomes (PROs).*
- *To evaluate the effect of TAK-062 on celiac serology titers.*
- *To evaluate the relationship between PROs and histological measures of intestinal injury.*
- *To evaluate the effect of simulated inadvertent gluten exposure (SIGE) on response to therapy with TAK-062.*
- *To evaluate urine gluten immunogenic peptide (GIP) detection as a means of monitoring gluten exposure and response to therapy.*

[redacted]

- [redacted]
- *To characterize the pharmacokinetic (PK) of TAK-062 in subjects with CeD attempting to maintain a GFD.*

1.2 Endpoints

1.2.1 Primary Endpoint

The primary endpoint is:

- *Change in CDSD gastrointestinal (GI) symptom severity score from baseline (Week -1) to Week 12.*

1.2.2 Secondary Endpoint

The secondary endpoint is:

- *Change in Vh:Cd from baseline (measured at Week -4) to Week 24.*

1.2.3 Safety Endpoints

Treatment-emergent adverse events (TEAEs), serious adverse events (SAEs), treatment-related TEAEs, electrocardiogram (ECG) findings, vital signs, laboratory parameters, and immunogenicity (posttreatment positive antidrug antibodies [ADA] in serum for TAK-062).

1.2.4 Exploratory/Additional Endpoints

Exploratory/additional endpoints (baseline assessments made [redacted] unless specified otherwise) are:

Impact of SIGE:

- *Differences in responses in CeD endpoints including PROs, celiac serology, histology [redacted] will be assessed between gluten-containing SIGE and gluten-free SIGE arms.*

PRO Outcomes:

- Proportion of Subjects Achieving a Change from Baseline in Weekly Average CDSD at Week 12 ≤ - 0.25 and ≤-0.44.
- *Change in CDSD GI symptom severity score from Week -8 and Week -1 to Week 24.*
- *Change in CDSD GI symptom severity score from Week -8 to Week 12.*
- *Patient Global Impression of Change (PGIC) at Week 0, Week 12, and Week 24.*
- *Percent of subjects reporting each symptom on the CeD Most Bothersome Symptom Questionnaire at each time point measured.*

- Change from baseline to Week 12 and Week 24 in percentage of symptom-free days over a 14-day period.
- Change from baseline to Week 12 and Week 24 in percentage of symptom-free days over a 7-day period.
- Proportion of subjects with ⩾55% of percentage of symptom-free days over a 14-day period at Week 24 or Early Termination visit.
- Proportion of subjects with ⩾55% of percentage of symptom-free days over a 7-day period at Week 24 or Early Termination Visit.
- Proportion of subjects with change from baseline to Week 24 in percentage of symptom-free days ⩾25% over a 14-day period.
- Change from baseline to Week 24 in average CDSD bowel movement frequency over a 14-day period.
- Change from baseline to Week 24 in average CDSD bowel movement frequency over a 7-day period.
- Change from baseline to Week 24 in percentage of symptom-free days over a 7-day period by PGIS category.
- Change from baseline to Week 24 in the following PROs (average over a 14-day period):
  - CDSD total score.
  - CDSD GI score.
  - CDSD non-stool GI score.
  - CDSD Diarrhea score.
  - CDSD Abdominal pain score.
  - CDSD Bloating score.
  - CDSD Nausea score.
- Change from baseline to Week 24 by PGIS category in the following PROs (average over a 7-day period):
  - CDSD total score.
  - CDSD GI score.
  - CDSD non-stool GI score.
  - CDSD Diarrhea score.
  - CDSD Abdominal pain score.
  - CDSD Bloating score.

  - CDSD Nausea score.
- *Change from Week -8 and Week -1 to Week 12 and Week 24 in the following PROs:*
  - *Weekly score of the most bothersome symptom as reported by the CDSD.*
  - *Severity rating or frequency of each item (symptom) of the CDSD.*
  - *Frequency of constipation based on the CDSD.*
  - *Proportion of minimal symptom days per week (defined as days with none or no more than 1 mild score of either bloating, abdominal pain, nausea, or diarrhea) as reported on the CDSD.*
- *Percentage of symptom-free days per week as reported on the CDSD.*
  - *Impact of Celiac Disease Symptoms Questionnaire (ICDSQ).*
  - *Patient Global Impression of Severity (PGIS).*
  - *Patient-reported Outcomes Measurement Information System (PROMIS)-Cognitive Function Instrument.*
  - *PROMIS-Fatigue Instrument.*
  - *Short Form 12-Item (SF-12) Health Survey.*
  - *EuroQol 5-Dimensions 5-Levels (EQ-5D-5L) health survey.*
  - *Work Productivity and Activity Impairment plus Classroom Impairment Questionnaire: Celiac Disease (WPAI+CIQ: CeD).*
  - *Celiac Disease Gastrointestinal Symptom Rating Scale (CeD-GSRS).*
  - *Celiac Disease Assessment Questionnaire (CDAQ).*
  - *Celiac disease quality of life (CD-QOL) questionnaire.*

Serology titers:

- *Change from baseline (Week -8) to* Week 12 and *Week 24 in celiac serology titers.*

Histological endpoints:

- Proportion of subjects achieving mucosal remission (Vh:Cd $\geqslant$3, $\geqslant$2.7 and >2.5) at Week 24.
- *Change from baseline (Week -4) to Week 24 in IEL counts.*
- *Change from baseline (Week -4) to Week 24 in Marsh-Oberhuber scores both qualitatively assessed and calculated from IEL and Vh:Cd results.*
- *Change from baseline (Week -4) to Week 24 in individual Vh and Cd measurements.*
- Proportion of subjects with an increase of ≥0, ≥0.1, ≥0.2, ≥0.3 and ≥0.4 in Vh:Cd from baseline to Week 24.

The following histological endpoints were not pre-specified in the study protocol but added in the SAP:

- Proportion of subjects with Vh:Cd ≥2.5 and IEL counts < 25 at Week 24.
- Mucosal response rate defined as the proportion of subjects with an increase in Vh:Cd of greater than 0.4 and a reduction in IEL counts of greater than 30% from Week -4 to Week 24.
- Proportion of subjects with an increase in Vh:Cd of greater than 0.25 and a reduction in IEL counts of greater than 15% from Week -4 to Week 24.
- Change from baseline (Week -4) to Week 24 in Composite score of Vh:Cd ratio and IEL (VCIEL) and Multi-parametric composite score of Vh:Cd and IEL (mpVCIEL).

PRO and histological endpoints:

- Proportion of subjects with an improvement in both CDSD GI symptom severity score and Vh:Cd from Week -4 to Week 24.
- Proportion of subjects with an improvement in both GSRS (from Week -8) and Vh:Cd (from Week -4) to Week 24.
- Proportion of subjects with Vh:Cd ⩾2.7 at Week 24 and ⩾55% of symptom-free days over a 14-day period at Week 24 for subgroup of subjects with baseline Vh:Cd ⩽2.3 and ⩽40% baseline percentage of symptom-free days.
- Proportion of subjects with change from baseline at Week 24 in Vh:Cd ⩾ 0.4 and change from baseline at Week 24 in percentage of symptom-free days ⩾ 25% over a 14-day period.

[redacted]

GIP endpoints:

- *Frequency of positive urine GIP tests in subjects across the treatment groups during screening, run in and treatment phases.*

PK endpoints:

- *Plasma concentration of TAK-062.*

1.3 Estimand(s)

Table 1.a Estimand Frameworks

Estimand: [Primary]

	Attributes				
Definition	**Treatment**	**Population**	**Variable (or Endpoint)**	**Strategy for Addressing Intercurrent Event**	**Population-Level Summary**
The primary estimand is the treatment effect of TAK-062 compared to Placebo at Week 12 in targeted patient population.	TAK-062 ([redacted]) versus placebo.	Subjects aged 18 and years and older, who are HLA-DQ2 and/or HLA-DQ8 positive, have biopsy-confirmed disease that is clinically active with ongoing GI symptoms and small intestinal villous atrophy on duodenal biopsy, have been attempting to maintain a GFD for ⩾12 months, and are experiencing inadvertent gluten exposure. *	Change in weekly average CDSD GI symptom severity score from baseline to Week 12.	Composite strategy will be used to address rescue medication use and treatment discontinuation due to lack of efficacy or treatment-related AEs. Treatment policy strategy will be used to address SIGE bar discontinuation and treatment discontinuation due to reasons other than lack of efficacy or treatment-related AEs.	Treatment difference (TAK-062 – placebo) in the mean change from baseline to Week 12 in weekly average CDSD GI symptom severity score, for each dose level.

** Adolescents (aged 12 to 17 years) and a maximum of 20% of adult subjects without a producible initial biopsy report confirming CeD may be enrolled if they meet the following inclusion criteria:*

a) Serology (IgA-tTg) at diagnosis or subsequent visit is at least 2 times the ULN.

b) Histology at screening biopsy at Week -4 must be consistent with Marsh-Oberhuber score of 2 or greater as read by a central pathologist.

2.0 STUDY DESIGN

Despite attempts to maintain a gluten-free diet (GFD), many individuals with celiac disease (CeD) continue to have symptoms and/or intestinal damage due to inadvertent gluten exposure. TAK-062 is a computationally designed enzyme intended to reduce or prevent symptoms and intestinal damage related to gluten exposure. This is a 2-cohort phase 2, multicenter, double-blind, randomized, placebo-controlled study to evaluate the efficacy and safety of TAK-062 for treatment of ongoing symptoms and small intestinal mucosal injury due to gluten exposure in subjects with CeD attempting to maintain a GFD. In addition, this study is designed to further evaluate the psychometric properties of the Celiac Disease Symptom Diary (CDSD) v2.1, [redacted]

[redacted], *and to provide data on the optimal histologic measure for assessment of response to treatment in CeD.*

Approximately 357 subjects with CeD and both ongoing symptoms and small intestine villous atrophy on duodenal biopsy, despite attempting to maintain a GFD for at least 12 months, will be enrolled and randomly assigned into 1 of 2 treatment groups in the Cohort 1proof of concept study and 1 of 6 treatment groups in the Cohort 2 dose-ranging study (see the Number of Subjects section below).

The interim analysis will be initiated when at least 75% of the subjects in Cohort 1 have either completed 12 weeks of treatment or have dropped out to determine inclusion of adolescents in Cohort 2. Cohort 1 data will be reviewed by an external independent data monitoring committee (IDMC) and shared with health authorities and independent ethics committees (IECs)/institutional review boards (IRBs) as required by region, before enrollment of adolescents in Cohort 2. If efficacy and safety data from Cohort 1 is considered acceptable, Cohort 2 will also include adolescents aged 12 to 17 years. Adolescents will receive only gluten-free simulated inadvertent gluten exposure (SIGE) bars.

Number of Subjects:

Cohort 1

Cohort 1 will consist of 120 adult subjects randomly assigned to receive 1 of 2 treatments:

1. TAK-062 placebo [redacted] + gluten-containing SIGE bar [redacted]: 60 subjects.

2. TAK-062 [redacted] + gluten-containing SIGE bar [redacted]: 60 subjects.

Cohort 2

Cohort 2 will consist of approximately 216 adult subjects to be enrolled and randomly assigned into 1 of 5 drug and SIGE treatment groups (Groups 1-5), and approximately 21 adolescent subjects to be enrolled and randomly assigned into Groups 4, 5 and 6 (adolescents only).

1. TAK-062 placebo [redacted] + gluten-containing SIGE bar [redacted]: 30 subjects (adults only).

2. TAK-062 [redacted] + gluten-containing SIGE bar [redacted]: 50 subjects (adults only).

3. TAK-062 [redacted] + gluten-containing SIGE bar [redacted]: 50 subjects (adults only).

4. TAK-062 placebo [redacted] + gluten-free SIGE bar [redacted]: 50 subjects (43 adults, 7 adolescents).

5. TAK-062 [redacted] + gluten-free SIGE bar [redacted]: 50 subjects (43 adults, 7 adolescents).

6. TAK-062 [redacted] + gluten-free SIGE bar [redacted]: 7 subjects (adolescents only).

This study will consist of 4 periods in each cohort: a 2- to 4-week screening period (Week -8 to Week -4), a 4-week single-blind placebo run-in (Week -4 to Day -1), a 24-week double-blind treatment period (Week 0 to Week 24), and a 4-week safety follow-up period. At the first visit (Week -8), subjects will consume 1 gluten-free SIGE bar to ensure no intolerance to any of the nongluten components of the bar. During the screening period (between Week -8 visit to Week -4

visit), subjects will complete the CDSD daily to confirm they have at least 1 ongoing CeD-related gastrointestinal (GI) symptom of moderate or greater severity on at least 3 days out of any consecutive 7-day period. The CeD-related symptom(s) may vary day by day as long as the severity of at least 1 symptom is moderate or greater. Eligible subjects will undergo an esophagogastroduodenoscopy (EGD) with duodenal biopsy [redacted] for the assessment of villous atrophy. In addition, a subset of approximately 40 subjects in total [redacted] During the screening period, urine samples will be collected once per week for gluten immunogenic peptide (GIP) to assess systemic exposure to gluten.

During the single-blind run-in period (Week -4 through Day -1), subjects will receive TAK-062 placebo [redacted] of the start of a meal and consume SIGE bar with or without gluten, [redacted] with a meal; [redacted]

Eligible subjects with villous atrophy on duodenal biopsy (defined as villous height to crypt depth ratio [Vh:Cd] <2.5 (<3 before protocol amendment 3) *by central pathology assessment) will be randomly assigned to 1 of 2 treatment groups in Cohort 1 or 1 of 6 treatment groups in Cohort 2 to receive double-blind treatment and gluten-containing SIGE or gluten-free SIGE (Cohort 2) for 24 weeks (Day 1 through Week 24). Initially, all subjects who meet symptom criteria and histologic criteria will be randomized. If more than 10% of planned enrollment in a cohort report a greater than 1 point improvement in Patient Global Impression of Severity during [redacted] run-in period (Week -2 through Day -1), further subjects showing this degree of improvement will be excluded from the cohort.*

Randomization will be stratified by celiac serologic status at Visit 1, defined as normal or elevated celiac serology (1 or more of the following: IgA-tissue transglutaminase (tTg) >1 times the upper limit of normal [ULN], IgA-deaminated gliadin peptide [DGP] >2 × ULN, or IgG-DGP >2 × ULN), mild to moderate (Vh:Cd 1.5 to <2.5) versus moderate to severe (Vh:Cd <1.5) histologic injury at baseline, and by use of proton pump inhibitors or histamine type 2 antagonists (yes or no) at Visit 1.

In Cohort 1, all subjects will consume gluten-containing SIGE bars [redacted] with a meal [redacted] In Cohort 2, adult subjects will receive either gluten-containing or gluten-free SIGE bars with a meal [redacted] throughout the treatment period. [redacted]

[redacted]. *It is expected that SIGE will result in stable, rather than worsening, symptoms and enteropathy in the TAK-062 placebo groups. If a subject has severe symptoms of diarrhea, nausea, vomiting, or abdominal spasm, that are related to their CeD symptoms, the use of the medications to treat these severe symptoms will be considered as rescue medications.*

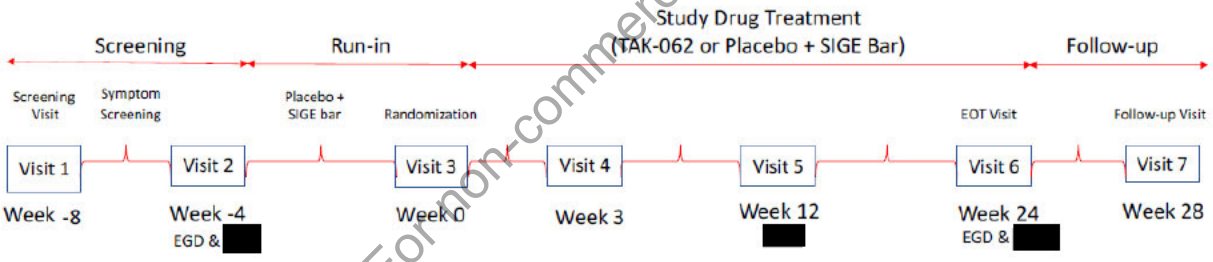
Subjects will report symptoms daily using the CDSD from Week -8 through Week 24.

EGD with biopsy will again be performed at the end of the 24-week treatment period (Week 24). [redacted]

Sparse blood samples to evaluate the pharmacokinetics (PK) of TAK-062 in plasma will be collected at Weeks -4, 0, 3, 12, 24, and 28. Approximately 10 adult subjects enrolled in Cohort 1 and 30 adult subjects enrolled in Cohort 2 will participate in the PK subgroup for slightly more intensive collections at Week 0 (randomization) and Week 3 visits.

There is a safety follow-up visit at Week 28, 4 weeks after the last date of treatment.

A diagram of the study design for Cohorts 1 and 2 is located below.



EGD: esophagogastroduodenoscopy; EOT: end of treatment; SIGE: simulated inadvertent gluten exposure; [redacted].

3.0 STATISTICAL HYPOTHESES AND DECISION RULES

3.1 Statistical Hypotheses

There are three hypotheses to be formally tested for the primary efficacy endpoint, change in CDSD GI symptom severity score from baseline (Week -1, the last week of the run-in period) to Week 12, in a fixed sequence. The null and alternative hypotheses for the 3 different doses against placebo are as follows:

H1	H_{10}: Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{TAK-062[redacted] +Gluten-containing SIGE bar}}$ = Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{Placebo + Gluten-containing SIGE bar}}$ versus

	H_{1A}: Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{TAK-062 [redacted] +Gluten-containing SIGE bar}}$ ≠ Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{Placebo + Gluten-containing SIGE bar}}$
H2	H_{20}: Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{TAK-062 [redacted] + Gluten-containing SIGE bar}}$ = Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{Placebo + Gluten SIGE bar}}$ versus H_{2A}: Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{TAK-062 [redacted] + Gluten-containing SIGE bar}}$ ≠ Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{Placebo + Gluten-containing SIGE bar}}$
H3	H_{30}: Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{TAK-062 [redacted] +Gluten-containing SIGE bar}}$ = Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{Placebo + Gluten-containing SIGE bar}}$ versus H_{3A}: Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{TAK-062 [redacted] + Gluten-containing SIGE bar}}$ ≠ Change from Baseline to Week 12 in CDSD GI symptom severity score $_{\text{Placebo + Gluten-containing SIGE bar}}$

3.2 Statistical Decision Rules

The statistical significance treatment effect will be tested against 2-sided alpha level of 0.05 following the pre-specified fixed-sequence order from high dose to low dose for primary efficacy endpoint.

1) TAK-062 [redacted] + gluten-containing SIGE bar versus placebo + gluten-containing SIGE bar.

2) TAK-062 [redacted] + gluten-containing SIGE bar versus placebo + gluten-containing SIGE bar.

3) TAK-062 [redacted] + gluten-containing SIGE bar versus placebo + gluten-containing SIGE bar.

3.3 Multiplicity Adjustment

To control the overall Type I error rate for the multiple comparisons performed for the primary efficacy endpoint, a fixed-sequence testing approach will be applied to the statistical testing. A hypothesis will only be tested if the statistical significance is achieved with the previous test (i.e., $p < 0.05$) in the sequence. Once a hypothesis fails to demonstrate the statistical significance (i.e., $p \geqslant 0.05$), all subsequent hypotheses will not be tested. However, nominal p-values and confidence intervals for the subsequent analyses will be provided and will be considered exploratory. If any treatment group is not opened or terminated early, the hypothesis testing of that treatment group will not be performed and the formal hypothesis testing will proceed with the next highest dose tested in the fixed sequence testing as pre-specified above. Multiplicity will

not be adjusted for secondary or exploratory efficacy endpoints. Nominal p-value will be presented.

4.0 SAMPLE-SIZE DETERMINATION

The planned sample size is approximately 357 subjects. For a 2-sample comparison of means, 0.5 standardized mean difference is regarded as a medium effect size. A sample size of 53 subjects per treatment group will provide 80% power to detect a standardized mean difference (mean difference/SD) of 0.55 between [redacted] TAK-062 and placebo in the change from baseline CDSD weekly score (assuming common SD) for a 2-sided statistical hypothesis test, using a 2-sample t-test and a 5% significance level. Assuming a dropout rate of approximately 12%, a sample size of 60 subjects per treatment group for the [redacted] and placebo arms in Cohort 1 will be the planned enrollment.

The planned enrollment for Cohort 2 is 30 and 50 subjects per treatment arm in the placebo and active arms ([redacted] TAK-062), respectively, in the gluten-containing SIGE arms; 50 subjects (43 adults and 7 adolescents) per arm in the [redacted] and placebo gluten-free SIGE bar arms; and 7 (adolescents only) in the [redacted] gluten-free SIGE bar arm.

For the combined placebo gluten-containing SIGE groups from Cohorts 1 and 2 versus [redacted] TAK-062, 86% power would be provided (based on the 2-sample t-test and the 5% significance level) to detect the standardized mean difference between groups (0.55) with the same 12% assumption for dropout rate. In addition, 82% power would be provided from the same assumptions for combined placebo gluten-containing SIGE group from Cohorts 1 and 2 versus [redacted] TAK-062 group. The analysis to detect the difference between the combined placebo gluten-containing SIGE group and the active group would yield a higher power.

5.0 ANALYSIS SETS

5.1 Full Analysis Set-SIGE

The Full Analysis Set (FAS)-SIGE will consist of all randomized subjects who are randomized to receive gluten-containing SIGE. Subjects will be analyzed according to the treatment they were randomized to receive.

The FAS-SIGE will be used for efficacy analysis for primary, secondary and exploratory efficacy endpoints.

5.2 Full Analysis Set-no SIGE

The FAS-no SIGE will consist of all randomized subjects who are randomized to gluten-free SIGE. Subjects will be analyzed according to the treatment they were randomized to receive.

The FAS-no SIGE will also be used for efficacy analysis for primary, secondary and exploratory efficacy endpoints.

5.3 Safety Analysis Set

The Safety analysis set (SAF) will consist of all randomized subjects who received at least 1 dose of study drug. Subjects will be analyzed according to actual treatment received.

The SAF will be used for safety analysis.

5.4 Safety Analysis Set-SIGE

The Safety analysis set (SAF)-SIGE will consist of all randomized subjects who received at least 1 dose of study drug and gluten-containing *SIGE*. Subjects will be analyzed according to actual treatment received.

The SAF-SIGE will be used for safety analysis for subjects who are randomized to receive gluten-containing SIGE.

5.5 Safety Analysis Set-no SIGE

The SAF-no SIGE will consist of all randomized subjects who received at least 1 dose of study drug and gluten-free SIGE. Subjects will be analyzed according to actual treatment received.

The SAF-no SIGE will be used for safety analysis for subjects who are randomized to gluten-free SIGE.

5.6 Per protocol analysis set-SIGE

The Per protocol analysis set (PPS)-SIGE will consist of all subjects in FAS-SIGE who do not violate the terms of the protocol in a way that would impact the study outcome. All decisions to exclude subjects for the PPS-SIGE will be made before the unblinding of the study.

The PPS will be used for the sensitivity analysis of the primary efficacy endpoint.

5.7 Pharmacokinetic Analysis Set

The Pharmacokinetic (PK) Analysis Set is defined as all randomized subjects who received at least 1 dose of study drug with at least 1 measured PK concentration. Subjects will be analyzed according to actual treatment received.

The PK analysis set will be used for PK analysis.

5.8 Immunogenicity Analysis Set

The Immunogenicity Analysis Set is defined as all randomized subjects who received any TAK-062 and have the baseline and at least 1 postbaseline immunogenicity sample assessment. Subjects will be analyzed according to actual treatment received.

The Immunogenicity Analysis Set will be used for immunogenicity analysis.

5.9 Randomized Set

The randomized set will consist of all subjects who were randomized. Subjects will be analyzed according to the treatment they were randomized to receive.

6.0 STATISTICAL ANALYSIS

6.1 General Considerations

Unless otherwise stated, baseline values are defined as the last observed value before the first dose of study medication.

All hypothesis tests and confidence intervals (CIs) will be 2-sided, and an alpha of 0.05 will be used for all statistical-testing, unless otherwise stated. All p-values reported will be 2-tailed and rounded to 3 decimal places prior to assessment of statistical significance.

Means and medians will be presented to 1 more decimal place than the recorded data. The standard deviations (SDs) will be presented to 2 more decimal places than the recorded data. CIs intervals will be presented using the same number of decimal places as the parameter estimate.

Where applicable, variables will be summarized descriptively by study visit. For the categorical variables, the counts and proportions of each possible value will be tabulated by treatment group. The denominator for the proportion will be based on the number of subjects who provided non-missing responses to the categorical variable. For continuous variables, the number of subjects with non-missing values, mean, median, SD, minimum, and maximum values will be tabulated.

A windowing convention will be used to determine the analysis value for a given study visit for observed data analyses. The details of window convention are described in Section 9.2.3.

Derived stratification factor will be used as the randomization stratification factor in the analyses unless specified otherwise.

The randomization stratification factor may be dropped from the analysis models for continuous variables and binary, ordinal and categorical variables (next section) if the number of subjects in the randomization factor are skewed in a factor due to the celiac serologic status, histologic injury, or use of proton pump inhibitors (PPIs) or histamine 2 antagonists (yes or no).

6.1.1 Handling of Treatment Misallocations

In the event of treatment (not SIGE) misallocations, the safety data will be analyzed using the treatment that the subject actually received, and the efficacy data will be analyzed using the treatment that the subject was randomized to receive regardless of the actual treatment received.

6.1.2 Analysis Approach for Continuous Variables

All continuous data will be summarized descriptively by analysis visit and treatment groups when appropriate. The analysis for the primary and secondary endpoints is described in detail in

Sections 6.5.1 and 6.5.2, respectively. The remainder of this section applies to exploratory analyses where statistical modeling is performed.

Longitudinal continuous data meeting the normal distribution assumption will be analyzed using a mixed-effects model for repeated measures (MMRM) based on the observed data only. The MMRM will include fixed effects for the TAK-062 treatment group, time, and treatment-by-time interaction, baseline randomization stratification variables and baseline outcome variables as covariates, and subject as a random effect. For longitudinal outcome measured daily or weekly, the time variable will be week. For longitudinal outcome measured at scheduled home or clinic visits, the time variable will be visit. An unstructured (co)variance structure will be used to model the within-subject errors. If the model fails to converge, another covariance structure such as autoregressive (1) or the compound symmetry covariance matrix may be considered in the order specified. Missing data will be left missing in the data.

Continuous data meeting the normal distribution assumption with only one post-baseline value will be analyzed using the ANCOVA model with treatment group as the independent variable, and baseline randomization stratification variables and corresponding baseline outcome variable as covariates. As a sensitivity analysis when the proportion of missing data is between 5% and 40%, inclusive, missing data can be imputed using multiple imputation method. The imputation models will include treatment group and baseline stratification variables.

Continuous efficacy data not meeting the normal distribution assumption will first be transformed such that a normal distribution can be assumed. The data analysis and missing data handling described above can then be applied to the transformed data. When data transformation is not feasible, the data can be analyzed using Wilcoxon rank-sum test. The Hodges-Lehmann estimator of location shift and the associated 95% CI will be presented for the treatment median difference between each TAK-062 dose arm and placebo. When a sensitivity analysis is necessary, tipping point analysis will be performed.

6.1.3 Analysis Approach for Binary, Ordinal and Categorical Variables

All binary, ordinal and categorical variables will be summarized descriptively by analysis visit and treatment groups when appropriate.

Longitudinal binary, ordinal and categorical outcome will be analyzed using generalized linear mixed model. The model will include fixed effects for the TAK-062 treatment group, time, and treatment-by-time interaction, baseline randomization stratification variables and baseline outcome variables as covariates, and subject as a random effect. For longitudinal outcome measured daily or weekly, the time variable will be week. For longitudinal outcome measured at scheduled home or clinic visits, the time variable will be visit. An unstructured (co)variance structure will be used to model the within-subject errors. If the model fails to converge, another covariance structure such as autoregressive (1) or the compound symmetry covariance matrix may be considered in the order specified. Missing data will be left missing in the data. Missing endpoint values will be adjusted for in the mixed model.

Binary, ordinal and categorical outcome with only one post-baseline measurement will be analyzed using logistic regression with treatment group as the independent variable, and baseline

randomization stratification variables, and baseline outcome variable, where applicable, as covariates. When a sensitivity analysis is necessary and the proportion of missing data is between 5% and 40%, inclusive, missing responses can be imputed using multiple imputation by fully conditional specification method.

6.2 Disposition of Subjects

General study information will be provided, including the date of first subject signing Informed Consent Form (ICF), date of first/last study drug dose, date of last subject's last visit/contact, date of last subject's last sample collection date for primary endpoint, Medical Dictionary for Drug Regulatory Activities (MedDRA) Version, World Health Organization Drug Dictionary (WHODrug) Version, and SAS Version used for creating the datasets.

Subject disposition will be summarized for each treatment group and overall, and will include the following:

- Summary of all screened subjects, including the number of screened subjects, number of subjects eligible/not eligible for randomization and primary reason for ineligibility for randomization.
- Summary of subjects randomized but not treated, completing study drug and completing study, and disposition of subjects based on the reasons for discontinuation of treatment and for failing to complete the study using the randomized set.
- When calculating percentages for the reasons for not being treated, the total number of subjects not treated will be used as the denominator. When calculating percentages for the reasons for discontinuation, the total number of subjects who prematurely discontinued will be used as the denominator.
- Summary of study subjects randomized by geographic region and site will be provided using the randomized set.
- Summary of subjects for each pre-defined population using the randomized set: FAS, FAS-no SIGE, SAF, SAF-SIGE, SAF-no SIGE, PPS-SIGE, PK analysis set and Immunogenicity analysis set. Reasons for exclusion from the FAS, FAS-no SIGE, SAF, SAF-SIGE, SAF-no SIGE, PPS-SIGE, PK analysis set, and Immunogenicity analysis set will be summarized and listed.

Significant protocol deviations will be summarized based on the randomized set.

Supporting data listings for randomization scheme, screen failures, subject disposition, and significant protocol deviations will be provided.

6.3 Demographic and Other Baseline Characteristics

6.3.1 Demographics

Descriptive summaries of demographic characteristics will be presented by treatment group and overall for FAS-SIGE, FAS-no SIGE, and SAF.

The following demographic characteristics will be summarized in the following order in the tables: age (years), age (categorical, <18, 18-<40 and >=40; adults and adolescents), sex, ethnicity, race, weight (kg), height (cm), and body mass index (BMI) (kg/m^2). Individual demographic characteristics will be listed.

6.3.2 Medical History and Concurrent Medical Conditions

Medical history and concurrent medical conditions will be coded using the MedDRA (Version 23 or higher). Medical history includes all medical conditions with a stop date prior to informed consent. A concurrent medical condition is a condition which is present at signing of informed consent.

Medical history and concurrent medical conditions will be summarized by system organ class (SOC), preferred term (PT) and by treatment group and overall using the SAF. The summary table 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 class even if he/she has multiple conditions/symptoms.

Individual subject medical history and concurrent medical condition data will be listed.

6.3.3 Baseline Characteristics

Descriptive summaries of baseline characteristics including stratification factors will be presented by treatment group and overall for FAS-SIGE, FAS-no SIGE, and SAF.

Stratification factors based on IRT data will be summarized along with the derived stratification factor based on lab report and collected concomitant medications in EDC. Any discrepancies will be presented in the data listings. Stratification factors are derived as follows.

- Celiac serologic status, defined as normal or elevated celiac serology (1 or more of the following: IgA-tissue transglutaminase (tTg) > 1× the upper limit of normal [ULN], IgA–deaminated gliadin peptide [DGP] > 2×ULN, or IgGDGP > 2×ULN) where ULN=7
- Histologic injury (at baseline): mild to moderate (Vh:Cd 1.5 to <2.5) versus moderate to severe (Vh:Cd <1.5)
- Use of proton pump inhibitors (PPIs) or histamine 2 antagonists (yes or no) at Visit 1

The following baseline celiac disease characteristics, including, but not limited to, HLA typing, celiac serologic titer, duration of CeD, method of diagnosis and associated findings, pathology findings at diagnostic endoscopy, presence of extraintestinal symptoms and anemia, reasons for diagnosis, celiac serologic status, histology injury status at screening (Vh:Cd) and PPI or histamine 2 antagonists use, gluten-free diet status and duration will be summarized. Individual baseline disease characteristics will be listed.

6.4 Medication History and Concomitant Medications

6.4.1 Prior Medications

Prior medications are defined as mediations that started and stopped prior to the first treatment administration. Prior medications will be coded using the latest version of the WHODrug Dictionary and summarized using the SAF.

Prior medications will be summarized by 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. If a subject reports taking 2 drugs belonging to the same class, he/she will only be counted once within that class.

Data listing(s) for prior medications will be provided.

6.4.2 Concomitant Medications

Concomitant medications are defined as medications that started prior to and were ongoing at initiation of first study treatment and those that started after initiation of first study treatment. The same analysis methods described in Section 6.4.1 will be conducted for concomitant medications.

6.5 Efficacy Analysis

6.5.1 Primary Endpoint Analysis

The primary endpoint is the change from baseline to Week 12 in CDSD GI symptom severity score.

6.5.1.1 Derivation of Endpoint

The CDSD v2.1 is a 5-item self-administered questionnaire that evaluates the severity and frequency of CeD symptoms on a daily basis. Severity is measured for the following CeD symptoms: diarrhea, abdominal pain, bloating, nausea, and tiredness. Symptom severity is evaluated using 5-point Likert-type scale ((0) none, (1) mild, (2) moderate, (3) severe, and (4) very severe). A supplemental questionnaire administered with the CDSD will measure the frequency of bowel movements, vomiting and diarrhea. The CDSD evaluates both severity and frequency of these symptoms over the previous 24 hours. The frequency of bowel movements, diarrhea, and vomiting is the number of instances of each symptom in the previous 24 hours. It will be administered as part of the daily diary for the duration of the study, including Screening, Run-in and treatment periods.

The weekly CDSD GI symptom severity score is an average of the daily GI symptom severity scores during the week. The daily GI symptom severity score is the average of the severity score for diarrhea, abdominal pain, bloating and nausea, ranging from 0 (no symptoms) to 4 (very severe). A lower score means less symptom severity. If any of the symptom severity score is missing, then the daily GI symptom severity score will set to missing. *Weekly CDSD scores will*

be calculated where the CDSD is completed on at least 4 of the 7 days. When the daily diary is completed on less than 4 of the 7 days, the weekly CDSD score will be considered missing in the primary analysis.

In the derivations described below, the GI symptom severity scores will be used in the primary analyses, the overall scores will be used in the supplementary analyses and the individual severity scores will be used in the exploratory analyses. The CDSD daily overall score will be calculated as the mean of the 5-point Likert-type scales of symptom severity in diarrhea, abdominal pain, bloating, nausea, and tiredness. The CDSD GI-only daily score is the mean of the scores for diarrhea, abdominal pain, bloating, and nausea. For CDSD overall or GI-only daily score, if one item has a missing value then the score will be set to missing for the day. Refer to Table 6.a for further details.

Table 6.a CDSD v2.1

CDSD v2.1 Question	Response	Include in Daily CDSD Overall Score?	Include in Daily GI symptom severity score?
1. During the past 24 hours, how severe was your diarrhea at its worst?	0=None 1=Mild 2=Moderate 3=Severe 4=Very severe	Yes	Yes
2. During the past 24 hours, how severe was your abdominal (belly) pain at its worst?	0=None 1=Mild 2=Moderate 3=Severe 4=Very severe	Yes	Yes
3. During the past 24 hours, how severe was your bloating (feeling as if you need to loosen your clothes) at its worst?	0=None 1=Mild 2=Moderate 3=Severe 4=Very severe	Yes	Yes
4. During the past 24 hours, how severe was your nausea (feeling as if you were going to vomit or throw up) at its worst?	0=None 1=Mild 2=Moderate 3=Severe 4=Very severe	Yes	Yes
5. During the past 24 hours, how severe was your tiredness at its worst?	0=None 1=Mild 2=Moderate 3=Severe	Yes	No

CDSD v2.1 Question	Response	Include in Daily CDSD Overall Score?	Include in Daily GI symptom severity score?
	4=Very severe		
6. How many times did you vomit (throw up) in the past 24 hours? (Please count each time you vomited rather than the number of trips to the bathroom).	Enter count.	No	No
7. How many bowel movements (poops) did you have in the past 24 hours?	Enter count.	No	No
8. How many of those bowel movements (poops) looked like type 6 or 7 in the picture below?	Enter count.	No	No

Day 1 will be defined as the date of first study treatment administration. Day -1 will be defined as the date of the last SIGE bar administration (as recorded in ePRO) and before the first dose of study medication based on the data as recorded in EDC. When the first dose is not recorded in EDC, the date of the first dose on or after the day of randomization in the e-diary will be considered Day 1.

Study day after the date of first dose of treatment will be calculated relative to Day 1 as: date of assessment/event – date of first dosing + 1. Study day prior to the first dosing will be calculated relative to Day -1 as: date of assessment/event – date of Day -1 – 1.

The baseline CDSD score will be the weekly average score of last week of 2 weeks when the subject is on gluten-containing SIGE in the run-in period and the CDSD is completed on at least 4 of the 7 days. When the daily diary is completed on less than 4 of the 7 days, first week data of 2 weeks when the subject is on gluten-containing SIGE [redacted] will be included in the average calculation while subject is on gluten-containing SIGE.

The analysis start date of run-in will be defined as min (date of first placebo IP in dosing diary, max [date of EGD, date of first SIGE bar administration at Visit 2]) as some subjects do not start SIGE bar until EGD is conducted. As subjects consume gluten-free SIGE bar during the first two weeks of run-in period, the first two weeks after the analysis start date of run-in will be excluded from the baseline calculation. Subject will be assumed to be on gluten-containing SIGE bar starting from the 3rd week of run-in period until Day -1.

Post-baseline weekly CDSD scores will be calculated based on analysis days where the CDSD is completed on at least 4 of the 7 days. When the daily diary is completed on less than 4 of the 7 days, the weekly CDSD score will be considered missing in the primary analysis. Any diary entered after the End of Study will be excluded from the summary. In the case that the diary is entered more than once on the same day, the last result based on time of entry will be used in the analysis.

The change from baseline to Weeks 12 in weekly CDSD GI symptom severity score will be calculated as

Change from baseline = weekly CDSD GI symptom severity score at Week 12 – CDSD GI symptom severity score at baseline,

6.5.1.2 Main Analytical Approach

The primary endpoint will be analyzed with a mixed model for repeated measures (MMRM) approach using the appropriate contrast at Week 12. The model will be based on all available weekly CDSD GI symptom severity scores from baseline through Week 12. The model will include treatment group, week, treatment-by-week interaction, and the randomization stratification factors as fixed effects and baseline CDSD GI symptom severity score as covariates, and subject as a random effect. An unstructured covariance structure will be used to model the within-subject errors. Restricted Maximum Likelihood (REML) will be used to fit the model and Kenwood-Roger method will be used to calculate degrees of freedom. If the model does not converge, additional covariance structure including autoregressive and compound symmetry will be used in the order stated. MMRM sample codes are provided in Section 9.4.

Two-sided tests comparing each of the TAK-062 dose groups with placebo at Week 12 will be conducted. The corresponding p-values, least squares (LS) means of treatment differences and 2-sided 95% CI will be reported. The MMRM assumes data is missing at random.

Multiplicity control: To control the overall type I error rate for the primary endpoint, a fixed-sequence testing procedure will be used for the comparison of 3 doses of TAK-062 versus placebo. Specifically, the primary endpoint tested will be conducted in the following order:

i. First, [redacted] TAK-062 versus placebo for subjects taking gluten-containing SIGE in Cohort 1.

ii. If the test in Step i) is statistically significant, then perform the superiority test of [redacted] TAK-062 versus combined placebo from Cohorts 1 and 2 for subjects taking gluten-containing SIGE.

iii. If the test in Step ii) is statistically significant, then perform the superiority test of [redacted] TAK-062 versus combined placebo from Cohorts 1 and 2 for subjects taking gluten-containing SIGE.

In steps ii) and iii), in the comparisons of [redacted] TAK-062 versus combined placebo and [redacted] TAK-062 versus combined placebo in the gluten-containing SIGE group, approximately 20 subjects last enrolled in the placebo arm in Cohort 1 from the same study sites as in Cohort 2 active treatment arms with gluten-containing SIGE will be included in the placebo arm.

6.5.1.3 Handling of Intercurrent Event

A composite strategy will be used to address rescue medication use and treatment discontinuation due to lack of efficacy or treatment-related AEs.

For subjects who received rescue medications for CeD related symptoms, the severity of the treated symptom(s) in CDSD on the days when the rescue medications are used will be set to Very Severe. For subjects who discontinue treatment due to lack of efficacy or treatment-related AEs, the Week 12 CDSD GI severity score will be set to the baseline weekly CDSD GI severity score. Treatment policy strategy will be used to address SIGE bar discontinuation (ie, the recorded CDSD GI symptom severity score will be used regardless of whether a subject has discontinued SIGE) and treatment discontinuation due to reasons other than lack of efficacy or treatment-related AEs.

6.5.1.4 Sensitivity Analysis

The following sensitivity analyses will be performed for the primary efficacy endpoint:

- The primary efficacy endpoint analysis will be repeated using the IRT randomization stratification factors as the fixed effects in the MMRM. Missing data will not be imputed.
- The primary efficacy endpoint analysis will be repeated without baseline randomization stratification factors as the fixed effects in the MMRM. Missing data will not be imputed.
- The primary efficacy endpoint analysis will be repeated using the PPS-SIGE. Missing data will not be imputed.
- The primary efficacy endpoint analysis will be repeated using treatment policy strategy *(ie, the recorded CDSD GI symptom severity score will be used)* to address all occurrences of treatment discontinuation due to any reason for all subjects. Missing data will not be imputed.
- The missing not at random (MNAR) will be examined using control-based pattern mixture model. Fully Conditional Specification (FCS) was used for multiple imputation by PROC MI FCS REG. The imputation models will include baseline stratification variables and previous measures. The imputation model for the missing values in the treatment arm is constructed from the observed data in the control (placebo) arm. The same model will be used to impute missing values in the control arm. The estimated treatment effect from the imputed datasets will then be combined to generate the statistical inference based on MNAR. Multiple imputation sample codes are provided in Section 9.4. The model will be constructed in 3 steps:
    - Step 1 (Imputations): A total of 1,000 datasets will be generated with imputed values, based on the control group data.
    - Step 2 (Analysis of complete data sets): The primary endpoint will be analyzed for each of the 1000 complete imputed data sets using an ANCOVA with treatment group and baseline strata as factors, and the baseline value as a covariate.
    - Step 3 (Inference): The LS mean estimates, associated standard error, 95% CI and p value will be aggregated using Rubin's rule via PROC MIANALYZE procedure.

- The examination of missing not at random (MNAR) using control-based imputation will be repeated. The primary endpoint will be analyzed with imputed data using a MMRM with treatment group and baseline strata as factors, and the baseline value as a covariate.
- The primary efficacy endpoint analysis will be repeated. A composite strategy will be used to address SIGE bar discontinuation. For subjects with SIGE bar discontinuation, the Week 12 CDSD GI severity score will be set to the baseline weekly CDSD GI severity score.
- The primary efficacy endpoint analysis will be repeated. A composite strategy will be used to address treatment discontinuation due CeD-related GI symptom worsening. For subjects with early termination due to CeD-related GI symptom worsening, the Week 12 CDSD GI severity score will be set to the baseline weekly CDSD GI severity score.
- The primary efficacy endpoint analysis will be repeated. A composite strategy will be used to address treatment discontinuation due CeD-related GI symptom worsening. For subjects with early termination due to CeD-related GI symptom worsening, the Week 12 CDSD GI severity score will be set to the worst CDSD GI severity score.
- The primary efficacy endpoint analysis will be repeated. For subjects who experience ICEs as described in Section 6.5.13, data following the ICEs will be imputed using the last observation carried forward (LOCF) method, based on the last available observation prior to the ICE.
- Additional sensitivity analyses, such as applying different baseline derivation rules, will be conducted as needed.

6.5.1.5 Supplementary Analyses

Supplementary analyses for the primary estimand will include the following:

- *Estimand with the target population of subjects who satisfied study entry criteria and are on pre-enrollment diet. The primary endpoint will be analyzed based on FAS–no SIGE. The endpoint, strategy for the ICEs and population-level summary will be the same as the main estimand. Nominal p-values will be calculated. No hypothesis testing will be performed.*
- *Estimand with the target population excluding subjects who report >1 point improvement in PGIS during [redacted] run-in period. The supplementary analysis will be performed for FAS-SIGE and FAS–no SIGE separately. The endpoint definition, strategy for the ICEs and population-level summary will be the same as the main estimand. Nominal p-values will be calculated. No hypothesis testing will be performed.*
- *Estimand with the weekly average CDSD-overall (GI + Tiredness) symptom severity score based on daily CDSD symptom severity score that is the average of the severity score for diarrhea, abdominal pain, bloating, nausea, and tiredness, ranging from 0 (no symptoms) to 4 (very severe). The population, treatments, strategy for the ICEs and population-level summary will be the same as the main estimand. The analysis will be*

performed based on FAS-SIGE and FAS–no SIGE separately. Nominal p-values will be calculated. No hypothesis testing will be performed.

- *Estimand using the area under the curve (AUC) of weekly CDSD score from Day 1 through Week 12 as the endpoint. The AUC will be a summation of daily scores. Missing values due to treatment discontinuation will be imputed by last observed carried forward (LOCF). The AUC will be analyzed using analysis of covariance with treatment group as an effect and baseline weekly CDSD score as covariance to compare the treatment difference with placebo and 2-sided 95% CI will be presented for each TAK-062 dose group. Nominal p-values will be calculated. No hypothesis testing will be performed.*

  AUC will be calculated as below:

  AUC = Sum of daily score from Day 1 to Week 12,

  Where the missing daily score is imputed by LOCF method.

- *Estimand using change in CDSD symptom score from Week -5 (before study intervention) to Week 12 as the endpoint. The analysis method will be the same as the main estimand. Nominal p-values will be calculated. No hypothesis testing will be performed.*

  Week -5 will be defined as the week before the run-in date.

6.5.2 Secondary Endpoint Analysis

The secondary endpoint is change in Vh:Cd ratio on duodenal biopsy from baseline (measured at Week -4) to Week 24. The Vh:Cd ratio represents mucosal architectural changes and a lower Vh:Cd ratio indicates more severe intestinal injury characterized by a flattening of the mucosa.

6.5.2.1 Derivation of Endpoint

Tissue samples to evaluate Vh:Cd will be collected through duodenal biopsies at Week -4 (Visit 2) and at Week 24 (Visit 6) or early termination. The change from baseline in Vh:Cd will be calculated as

change from baseline = value at Week 24/ET – value at baseline,

where baseline is defined as the result of last observed value before the first dose of study medication, i.e., the result at Week -4.

6.5.2.2 Main Analytical Approach

The main estimand for the secondary objective assess the treatment effect of TAK-062 compared with placebo at Week 24 in target population based on the following variable/endpoint:

- *Change in Vh:Cd from baseline (measured at Week -4) to Week 24.*

The target population and treatments are the same as for the primary estimand in Section 0. Composite strategy will be used to address the ICEs of treatment discontinuation due to lack of efficacy or treatment-related AEs; treatment policy strategy for the ICEs of SIGE bar discontinuation or rescue medication use (i.e., Vh:Cd measured at Week 24 will be used

regardless of SIGE bar discontinuation or rescue medication use) and other reasons for discontinuations.

The corresponding population-level summaries are:

- Treatment difference (TAK-062 – placebo) in the change from baseline to Week 24 in average Vh:Cd on duodenal biopsy.

The analysis will be conducted using ANCOVA. The model will include treatment group and the randomization stratification factors as fixed effects and baseline values as covariate. Missing values will be regarded as missing at random. Missing data will be handled using multiple imputation. The missing values will be imputed by treatment group, stratification factors, gender and baseline Vh:Cd via PROC MI FCS REG. 1000 complete datasets will be computed. Each of the 1000 imputed datasets will be analyzed using ANCOVA. The LS mean estimates, associated standard error, 95% CI and p value will be aggregated using Rubin's rule via PROC MIANALYZE procedure. *The LS means and 2-sided 95% CI will be provided for the Week 24 treatment difference between the* [redacted] *TAK-062 dose groups versus placebo, respectively. Nominal 2-sided p-values will be provided for the comparison of* [redacted] *TAK-062 versus placebo,* [redacted] *TAK-062 versus placebo and* [redacted] *TAK-062 versus placebo, respectively for both FAS-SIGE and FAS-no SIGE.* Multiple imputation sample codes are provided in Section 9.4.

6.5.2.3 Sensitivity Analysis

The following sensitivity analyses will be performed for the secondary endpoint:

- The analyses will be repeated using the PPS-SIGE. Missing data will not be imputed.
- The analyses will be conducted without missing data imputation, i.e., missing as missing.

6.5.2.4 Supplementary Analyses

Supplementary analyses will be planned as follows:

Supplementary analyses based on subjects who satisfied study entry criteria and are on pre-enrollment diet will be performed. The corresponding endpoints will be analyzed based on FAS-no SIGE.

Supplementary analyses based on the target population excluding subjects who report >1 point improvement in PGIS during [redacted] *run-in period will also be performed. The analysis will be performed for FAS-SIGE and FAS-no SIGE separately.*

For all the above supplementary analyses, the endpoint definitions, strategy for the ICEs and population-level summary will be the same as for the main estimand. Nominal p-values will be calculated. No hypothesis testing will be performed.

6.5.2.5 Exploratory Analysis

As an exploratory analysis, the endpoint of change from baseline (Week -4) in Vh:Cd measurements at Week 24/ET will also be analyzed by modeling based on combined FAS-SIGE

and FAS-no SIGE excluding the [redacted] TAK-062 [redacted] group, where the effects of treatment group, SIGE and their interaction effects will be estimated in the same model.

The analysis on Vh:Cd will be performed similarly for the changes in the individual Vh and Cd measurements as exploratory analyses.

6.5.3 Exploratory/additional Endpoints Analysis

Exploratory/additional endpoints will be summarized descriptively by analysis visit and treatment groups in both FAS-SIGE set and FAS-no SIGE set when appropriate. MMRM, Logistic regression model and ANCOVA will be applied when appropriate.

Exploratory/additional endpoints are:

Impact of SIGE:

- Differences in responses in CeD endpoints including PROs, celiac serology, and histology measures will be compared between gluten-containing SIGE and gluten-free SIGE arms.

  Analysis will be based on subjects in Cohort 1 and the gluten-free SIGE arms (TAK-062 [redacted] and Placebo) in Cohort 2. Descriptive statistics will be provided on the change-from-baseline values for Weeks 12 (where applicable) and 24 for PROs, celiac serology, and histology measures for Control and TAK-062[redacted]. ANCOVA analysis will be performed with Treatment and SIGE as independent variables, and baseline randomization stratification variables and corresponding baseline outcome variable as covariates.

PRO Outcomes:

- Proportion of Subjects Achieving a Change from Baseline in Weekly Average CDSD at Week 12 ≤ - 0.25 and ≤ -0.44

  The weekly average CDSD score experiencing ICEs will be handled following rule 6.5.1.3. Missing values not related to ICEs will be assumed missing at random and addressed using multiple imputation (MI) with the PROC MI FCS REG method. The imputation model will include treatment group, stratification factors, and previous measures as covariates. A total of 1,000 complete datasets will be generated.

  The response variable will be dichotomized based on the imputed CDSD weekly average score. Response rates for each treatment group and the treatment difference will be analyzed across the imputed datasets. The point estimate, standard error, confidence intervals, and p-values will be aggregated using Rubin's rule, implemented with PROC MIANALYZE.

- Change in weekly average CDSD GI symptom severity score from Week -8 to Week 24

  Change in weekly average CDSD GI symptom severity score from Week -8 to Week 24 will be summarized by descriptive statistics for each treatment arm.

- Change in weekly average CDSD GI symptom severity score from Week -8 to Week 12.

  Change in weekly average CDSD GI symptom severity score from Week -8 to Week 12 will be summarized by descriptive statistics for each treatment arm.

- Patient Global Impression of Change (PGIC) at Week 0, Week 12, and Week 24.

  PGIC will be summarized as a categorical variable for Week 0, Week 12 and Week 24 by treatment arm.

- Percent of subjects reporting each symptom on the CeD Most Bothersome Symptom Questionnaire at each time point measured.

  The number and percent of subjects reporting each symptom as the most bothersome symptom will be summarized by visit and treatment arm.

- Change from baseline to Week 12 and Week 24 in percentage of symptom-free days over a 14-day period.

  Symptom-free days will be counted based on analysis days where the daily diary is completed on at least 8 of the 14 days. When the daily diary is completed on less than 8 of the 14 days, the symptom-free day will be counted as missing. Any diary entered after the End of Study will be excluded from the summary. In the case that the diary is entered more than once on the same day, the last result based on time of entry will be used in the analysis. The same scenario will be applied on other additional endpoints defined over a 14-day period. The percentage of symptom-free days is defined as the number of symptom-free days over the number of non-missing daily diary days over a 14-day period.

- Change from baseline to Week 12 and Week 24 in percentage of symptom-free days over a 7-day period.

  Symptom-free days will be counted based on analysis days where the daily diary is completed on at least 4 of the 7 days. When the daily diary is completed on less than 4 of the 7 days, the symptom-free day will be counted as missing. Any diary entered after the End of Study will be excluded from the summary. In the case that the diary is entered more than once on the same day, the last result based on time of entry will be used in the analysis. The same scenario will be applied on other additional endpoints defined over a 7-day period. The percentage of symptom-free days is defined as the number of symptom-free days over the number of non-missing daily diary days over a 7-day period.

- Proportion of subjects with $\geqslant$55% of symptom-free days over a 14-day period at Week 24 or Early Termination Visit

  Subgroup analysis will be also performed in subjects with $\leqslant$40% baseline percentage of symptom-free days. Other percentage of symptom-free days thresholds will be explored when appropriate.

- Proportion of subjects with ⩾55% of symptom-free days over a 7-day period at Week 24 or Early Termination Visit.

  Subgroup analysis will be also performed in subjects with ⩽40% baseline percentage of symptom-free days. Other percentage of symptom-free days thresholds will be explored when appropriate.

- Proportion of subjects with change from baseline to Week 24 in percentage of symptom-free days ⩾25% over a 14-day period.

  Subgroup analysis will be also performed in subjects with ⩽40% baseline percentage of symptom-free days. Other percentage of symptom-free days thresholds will be explored when appropriate.

- Change from baseline to Week 24 in average CDSD bowel movement frequency over a 14-day period.

  Change from baseline at Week 24 in in average CDSD bowel movement frequency will be summarized according to the CDSD V2.1 Question 8 in Table 6.a.

- Change from baseline to Week 24 in average CDSD bowel movement frequency over a 7-day period.

  Change from baseline at Week 24 in in average CDSD bowel movement frequency will be summarized according to the CDSD V2.1 Question 8 in Table 6.a.

- Change from baseline to Week 24 in percentage of symptom-free days over a 7-day period by PGIS category.

- Change from baseline to Week 24 in the following PROs (average over a 14-day period):
  - CDSD total score.
  - CDSD GI score.
  - CDSD non-stool GI score.
  - CDSD Diarrhea score.
  - CDSD Abdominal pain score.
  - CDSD Bloating score.
  - CDSD Nausea score.

  CDSD total score, GI score, non-stool GI score, Diarrhea score, Abdominal pain score, Bloating score, and Nausea score will be calculated based on Table 6.a. Non-stool GI score includes Abdominal pain score, Bloating score and Nausea score. The analysis of change from baseline at Week 24 in average scores over a 14-day period will follow the methodology mentioned in Section 6.5.1.2.

- Change from baseline to Week 24 by PGIS category in the following PROs (average over a 7-day period):
  - CDSD total score.
  - CDSD GI score.
  - CDSD non-stool GI score.
  - CDSD Diarrhea score.
  - CDSD Abdominal pain score.
  - CDSD Bloating score.
  - CDSD Nausea score.

  CDSD total score, GI score, non-stool GI score, Diarrhea score, Abdominal pain score, Bloating score, and Nausea score will be calculated based on Table 6.a. Non-stool GI score includes Abdominal pain score, Bloating score and Nausea score. The analysis of change from baseline at Week 24 in average scores over a 7-day period will follow the methodology mentioned in Section 6.5.1.2.
- Change from Week -8 and Week -1 to Week 12 and Week 24 in the following PROs:

  Change from Week -8 to Week 12 and Week 24, and separately, from Week -1 to Week 12 and Week 24, will be summarized by Visit and treatment group for the following:
  - Weekly score of the most bothersome symptom as reported by the CDSD.
  - Severity rating or frequency of each item (symptom) of the CDSD.
  - Frequency of constipation based on the CDSD.
  - Proportion of minimal symptom days per week (defined as days with none or no more than 1 mild score of either bloating, abdominal pain, nausea, or diarrhea) as reported on the CDSD.
  - Proportion of symptom-free days (response of "none" for all items) per week as reported on the CDSD.
  - Impact of Celiac Disease Symptoms Questionnaire (ICDSQ).
  - Patient Global Impression of Severity (PGIS).
  - Patient-reported Outcomes Measurement Information System (PROMIS)-Cognitive Function Instrument.
  - PROMIS-Fatigue Instrument.
  - Short Form 12-Item (SF-12) Health Survey.
  - EuroQol 5-Dimensions 5-Levels (EQ-5D-5L) health survey.

- Work Productivity and Activity Impairment plus Classroom Impairment Questionnaire: Celiac Disease (WPAI+CIQ: CeD).
- Celiac Disease Gastrointestinal Symptom Rating Scale (CeD-GSRS).
- Celiac Disease Assessment Questionnaire (CDAQ).
- Celiac disease quality of life (CD-QOL) questionnaire.

Serology titers:

- Change from baseline (Week -8) to Week 12 and Week 24 in celiac serology titers.

  Change from Week -4 to Week 12 and Week 24 in celiac serology titers will be summarized by Visit and treatment group.

Histological endpoints:

The following will be summarized with descriptive statistics by treatment arm for Week 24 visit and early termination visit.

- Proportion of subjects achieving mucosal remission (Vh:Cd ≥3, ≥2.7 and >2.5) at Week 24.
- Change from baseline (Week -4) to Week 24 in IEL counts.
- Change from baseline (Week -4) to Week 24 in Marsh-Oberhuber scores, both qualitatively assessed and calculated from IEL and Vh:Cd results.
- Change from baseline (Week -4) to Week 24 in individual Vh and Cd measurements.
- Proportion of subjects with Vh:Cd ≥2.5 and IEL counts < 25 per 100 enterocytes at Week 24.
- Mucosal response rate defined as the proportion of subjects with an increase in Vh:Cd of greater than 0.4 and a reduction in IEL counts of greater than 30% from Week -4 to Week 24.
- Proportion of subjects with an increase in Vh:Cd of greater than 0.25 and a reduction in IEL counts of greater than 15% from Week -4 to Week 24.
- Change from baseline (Week -4) to Week 24 in Composite score of Vh:Cd ratio and IEL (VCIEL) and Multi-parametric composite score of Vh:Cd and IEL (mpVCIEL).
- Proportion of subjects with an increase of ≥0, ≥0.1, ≥0.2, ≥0.3 and ≥0.4 in Vh:Cd from baseline to Week 24

  Subgroup analysis will be also performed in subjects with baseline Vh:Cd ≥2.3, or by stratification factors.

PRO and histological endpoints:

The following will be summarized with descriptive statistics by treatment arm for Week 24 visit and early termination visit. Improvement for CDSD and GSRS is defined as any reduction in score. Improvement in Vh:Cd is defined as any increase in Vh:Cd.

- Proportion of subjects with an improvement in both weekly average CDSD GI symptom severity score and Vh:Cd from Week -4 to Week 24
- Proportion of subjects with an improvement in both GSRS (from Week -8) and Vh:Cd (from Week -4) to Week 24
- Proportion of subjects with Vh:Cd $\geqslant$2.7 at Week 24 and $\geqslant$55% of symptom-free days over a 14-day period at Week 24 for subgroup of subjects with baseline Vh:Cd $\leqslant$2.3 and baseline percentage of symptom-free days $\leqslant$40%.

  Other percentage of symptom-free days thresholds will be explored when appropriate.
- Proportion of subjects with change from baseline to Week 24 in Vh:Cd $\geqslant$ 0.4 and change from baseline to Week 24 in percentage of symptom-free days $\geqslant$ 25% over a 14-day period.

  Subgroup analysis will be also performed in subjects with baseline percentage of symptom-free days $\leqslant$40%. Other percentage of symptom-free days thresholds will be explored when appropriate.

[redacted]

GIP endpoints:

- Frequency of positive urine GIP tests in subjects in screening and then by week across the different treatment groups.

PK endpoints:

- Plasma concentration of TAK-062.
- PK parameters

The following PK parameters will be calculated based on actual sampling time when applicable.

- PK subgroup: Cmax, Tmax, AUC0-3, AUC0-t, CL/F, T1/2, and Vz/F

  Cmax: the maximum observed drug concentration in plasma

  Tmax: the time at which Cmax occurs

AUC0-3: area under the concentration time curve from zero to 3 h after TAK-062 administration

AUC0-t: area under the concentration time curve from zero to time of last quantifiable concentration

CL/F: plasma clearance

T1/2: terminal half life

Vz/F: apparent volume of distribution

- All other subjects: Cmax

6.5.3.1 Derivation of Exploratory Endpoints

Impact of SIGE:

The difference of CeD endpoints including PROs, celiac serology, histology [redacted] will be assessed.

PRO Outcomes:

All PROs will be administered on an electronic device. The CDSD will be administered at approximately the same time every day. All PROs should be done, before any other procedures on the day of the study visit or no more than 1 day before the study visit. There is no expected missing item responses as the PRO instrument was set up to ask the subjects to answer the question only when the previous items are answered. Therefore there is no missing handling in the derivation of PRO endpoints unless specified otherwise.

PGIC

The Celiac Disease PGIC is a 1-question instrument that evaluates the change from study start to the present time on patients' severity. Response options are measured using a 7-point Likert-type scale ((7) very much worse, (6) much worse, (5) a little worse, (4) no change, (3) a little improved, (2) much improved, (1) very much improved). The PGIC will be conducted at Week 0 (Visit 3), Week 12 (Visit 4), Week 24 (Visit 6) or Early termination.

The PGIC scores can be collapsed and categorized as below:

PGIC Score	Category
1 = very much improved	A = Improved
2 = much improved	A = Improved
3 = a little improved	B = Minimal – No Change
4 = no change	B = Minimal – No Change
5 = a little worse	B = Minimal – No Change

6 = much worse	C = Worse
7 = very much worse	C = Worse

Note: B or C are classified as "Not Improved" as dichotomized as "Improved" vs "Not Improved.

CeD Most Bothersome Symptom

The CeD Most Bothersome Symptom is a 1-question instrument designed to assess the subject's perception of which symptoms from those evaluated in the CDSD that the subject considers most bothersome. The CeD Most Bothersome Symptom will be conducted at Week -8 (Visit 1), Week 0 (Visit 3), Week 12 (Visit 4), Week 24 (Visit 6) or Early termination.

ICDSQ

The ICDSQ is a 14-question instrument that evaluates emotional wellbeing, physical activities, social activities, and daily activities over the past 7 days. Responses are measured using a 5-point Likert-type scale ((0) not at all, (1) a little, (2) moderately, (3) very much, (4) completely). The ICDSQ will be conducted at Week -8 (Visit 1), Week 0 (Visit 3), Week 12 (Visit 4), Week 24 (Visit 6) or Early termination.

The domain score will be calculated as the mean score of all items in each domain. The overall ICDSQ score will be calculated as the sum of all domain scores. A higher score implies more sever symptom.

The change from baseline in domain score and overall score will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result of last observed value before the first dose of study medication, i.e., the result at Week 0. The change from Week -8 to post-baseline values will also be analyzed, where Week -8 is defined as the first available result before treatment administration.

PGIS

PGIS is a 1-question instrument used as an anchor that evaluates the overall severity over the past 7 days. Response options are measured using a Likert-type scale ((0) no symptoms, (1) mild, (2) moderate, (3) severe, (4) very severe). For validation purposes, it will be performed at Weeks -8 and -7. The PGIS will also be conducted on the day of every visit, i.e., Week -8 (Visit 1), Week -4 (Visit 2), Week -2, Week 0 (Visit 3)/Day -1, Week 3 (Visit 4), Week 12 (Visit 5), Week 24 (Visit 6) or early termination.

The change from baseline in PGIS will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result of last observed value before the first dose of study medication, i.e., the result at Week 0. The change from Week -8 to post-baseline values will also be analyzed, where Week -8 is defined as the first available result before treatment administration.

PROMIS-Cognitive Function Instrument

PROMIS is a set of person-centered measures that evaluates and monitors physical, mental, and social health in adults and children. The PROMIS Short Form v2.0-Cognitive Function 6a is a subset of 6 items assessing patient-perceived cognitive deficits. Facets include mental acuity, concentration, verbal and nonverbal memory, verbal fluency, and perceived changes in these cognitive functions. The extent to which cognitive impairments interfere with daily functioning, whether other people observe cognitive impairments, and the impact of cognitive dysfunction on quality of life are also assessed. This instrument has a 7-day recall period, and responses are measured using a 5-point frequency scale ((5) never, (4) rarely, (3) sometimes, (2) often, (1) very often). For subjects aged 12 to 17 years, a modified version of the PROMIS Pediatric Short Form v1.0-Cognitive Function 7a: Pediatric Cognitive Function will be used. This instrument has 7 items on a 5-point response scale ((5) none of the time, (4) a little of the time, (3) some of the time, (2) most of the time, (1) all of the time). The original version has a 4-week recall that has been modified to 7 days.

The PROMIS-Cognitive Function Instrument will be conducted at Week -8 (Visit 1), Week 0 (Visit 3), Week 12 (Visit 4), Week 24 (Visit 6) or early termination.

Raw scores (sum of all scale scores) will be converted to T-scores, using scoring tables Table 6.b and Table 6.c based on subject age.

Table 6.b Cognitive Function Scoring Table for Adult Subjects

Adult v2.0 - Cognitive Function 6a		
Short Form Conversion Table		
Raw Score	T-Score	SE*
6	23.13	4.25
7	26.64	3.28
8	28.55	3.05
9	30.18	2.84
10	31.58	2.72
11	32.85	2.64
12	34.04	2.59
13	35.17	2.57
14	36.28	2.57
15	37.37	2.57
16	38.45	2.57
17	39.53	2.58
18	40.63	2.59
19	41.74	2.60
20	42.87	2.62
21	44.04	2.63
22	45.23	2.64
23	46.47	2.67
24	47.77	2.71
25	49.17	2.79
26	50.72	2.94
27	52.49	3.14
28	54.69	3.51
29	57.60	4.04
30	63.17	5.75
*SE= Standard Error on T-score metric		

Table 6.c Cognitive Function Scoring Table for Subjects Aged 12-17

Pediatric v1.0 - Cognitive Function 7a		
Short Form Conversion Table		
Raw Score	T-Score	SE*
7	24.01	3.98
8	27.66	2.72
9	29.47	2.42
10	30.9	2.23
11	32.11	2.11
12	33.18	2.04
13	34.18	1.99
14	35.11	1.97
15	36.01	1.95
16	36.89	1.95
17	37.76	1.94
18	38.62	1.95
19	39.47	1.95
20	40.33	1.95
21	41.19	1.95
22	42.07	1.96
23	42.96	1.97
24	43.88	1.99
25	44.83	2.01
26	45.82	2.03
27	46.84	2.04
28	47.90	2.06
29	49.02	2.08
30	50.22	2.11
31	51.54	2.17
32	53.02	2.30
33	54.79	2.57
34	57.26	3.13
35	63.09	5.40
*SE= Standard Error on T-score metric		

The change from baseline in PROMIS-Cognitive Function Instrument t-score will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result at of last observed value before the first dose of study medication, i.e., the result at Week 0. The change from Week -8 to post-baseline values will also be analyzed, where Week -8 is defined as the first available result before treatment administration.

PROMIS-Fatigue Instrument

PROMIS is a set of person-centered measures that evaluates and monitors physical, mental, and social health in adults and children. The PROMIS Item Bank v1.0-Fatigue-Short Form 13a is a battery of 13 questions for adults aged 18 years or older, which are included in the FACIT-Fatigue. This instrument assesses a range of self-reported symptoms, from mild subjective feelings of tiredness to an overwhelming, debilitating, and sustained sense of exhaustion that likely decreases one's ability to execute daily activities and function normally in family or social roles.

Fatigue is divided into the experience of fatigue (frequency, duration, and intensity) and the impact of fatigue on physical, mental, and social activities. This instrument has a 7-day recall period, and responses are measured using a 5-point scale ((1) not at all, (2) a little, (3) somewhat, (4) quite a bit, (5) very much). The assigned numbers to the scale for 2 questions (AN5-I have energy and AN7-I am able to do my usual activities) are reversed from 5 to 1.

For subjects aged 12 to 17 years, the PROMIS Pediatric Item Bank v2.0-Fatigue-Short Form 10a: Pediatric Fatigue will be used. This instrument has 10 questions with a 7-day recall that are measured using a 5-point scale ((1) never, (2) almost never, (3) sometimes, (4) often, (5) almost always).

The PROMIS-Fatigue Instrument will be conducted at Week -8 (Visit 1), Week 0 (Visit 3), Week 12 (Visit 4), Week 24 (Visit 6) or early termination.

Raw scores (sum of all scale scores) will be converted to T-scores, using scoring tables Table 6.d and Table 6.e based on subject age.

Table 6.d Fatigue Scoring Table for Adult Subjects

Fatigue 13a (FACIT-Fatigue) - Adult v1.0 *Short Form Conversion Table*					
Raw Score	T-score	SE*	Raw Score	T-score	SE*
13	30.3	4.7	40	60.8	1.8
14	35.0	3.5	41	61.4	1.8
15	38.0	3.0	42	62.0	1.8
16	40.3	2.8	43	62.6	1.8
17	42.1	2.6	44	63.2	1.8
18	43.7	2.5	45	63.8	1.8
19	45.0	2.3	46	64.4	1.8
20	46.3	2.2	47	65.0	1.8
21	47.3	2.1	48	65.6	1.8
22	48.3	2.0	49	66.2	1.9
23	49.3	2.0	50	66.9	1.9
24	50.1	1.9	51	67.5	1.9
25	51.0	1.9	52	68.2	1.9
26	51.7	1.9	53	68.9	2.0
27	52.5	1.9	54	69.6	2.0
28	53.2	1.9	55	70.4	2.0
29	53.9	1.8	56	71.2	2.1
30	54.6	1.8	57	72.0	2.2
31	55.3	1.8	58	72.9	2.3
32	55.9	1.8	59	73.9	2.4
33	56.6	1.8	60	75.0	2.5
34	57.2	1.8	61	76.2	2.7
35	57.8	1.8	62	77.5	2.9
36	58.4	1.8	63	79.1	3.1
37	59.0	1.8	64	81.2	3.3
38	59.6	1.8	65	83.5	3.4
39	60.2	1.8			

*SE = Standard Error on T-score metric

Table 6.e Fatigue Scoring Table for Subjects Aged 12-17

Fatigue 10a - Pediatric v2.0					
Short Form Conversion Table					
Raw Score	T-Score	SE*	Raw Score	T-Score	SE*
10	30.3	5.5	31	60.6	3.3
11	34.3	4.7	32	61.6	3.3
12	36.9	4.4	33	62.6	3.3
13	39	4.1	34	63.6	3.3
14	40.9	3.9	35	64.6	3.3
15	42.5	3.8	36	65.6	3.3
16	44	3.7	37	66.7	3.3
17	45.4	3.6	38	67.7	3.3
18	46.7	3.5	39	68.7	3.3
19	47.9	3.5	40	69.8	3.3
20	49.1	3.4	41	70.9	3.3
21	50.2	3.4	42	72	3.4
22	51.3	3.4	43	73.2	3.4
23	52.4	3.4	44	74.4	3.4
24	53.5	3.4	45	75.7	3.5
25	54.5	3.4	46	77	3.6
26	55.6	3.4	47	78.5	3.6
27	56.6	3.4	48	80.2	3.7
28	57.6	3.4	49	82	3.7
29	58.6	3.3	50	84	3.5
30	59.6	3.3			
*SE= Standard Error on T-Score metric					

The change from baseline in PROMIS-Fatigue Instrument will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result at of last observed value before the first dose of study medication, i.e., the result at Week 0. The change from Week -8 to post-baseline values will also be analyzed, where Week -8 is defined as the first available result before treatment administration.

SF-12 v2

The SF-12 v2 is a self-administered, validated questionnaire designed to measure generic HRQOL. This 12-item questionnaire measures 8 domains, including: Physical Functioning, Role-physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-emotional, and Mental Health. Two summary scores can be calculated, the Physical Component Score (PCS), and the Mental Component Score (MCS). Higher scores indicate better health status.

The SF-12 v2 will be conducted at Week -8 (Visit 1), Week 0 (Visit 3), Week 12 (Visit 4), Week 24 (Visit 6) or early termination.

Raw scores range from 0 to 100 with higher scores indicating better health. Domain scores are calculated from raw scores such that domain scores have a mean of 50 and SD of 10. The PCS and MCS summary component scores also have mean of 50 and SD of 10 to allow comparisons with domain scores.

Scoring the SF-12 v2 is accomplished using T-score Based scoring software from Quality Metric Inc. (Lincoln, RI). T-score Based scoring is standardized across the SF family of adult tools using the means and standard deviations from the 2009 U.S. general population. The T-score Based scores in the U.S. general population have a mean of 50 and a standard deviation of 10. The Medical Outcomes Study (MOS) tools utilize 2009 t-scores.

T-score based scoring method scores the data in relation to U.S. general population t-scores. Therefore, all scores obtained that are below 50 can be interpreted as below the U.S. general population t-score and scores above 50 can be interpreted as above the U.S. general population t-score. Low values represent a poor health state and high values represent a good mental health. The change from baseline in SF-12 v2 domain/summary score will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result at of last observed value before the first dose of study medication, i.e., the result at Week 0. The change from Week -8 to post-baseline values will also be analyzed, where Week -8 is defined as the first available result before treatment administration.

EQ-5D-5L

The EQ-5D-5L questionnaire, developed by the EuroQol Research Foundation is a simple, valid, and reliable instrument used to measure general health-related quality of life (HRQOL) in subjects and includes 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Within each dimension, subjects choose 1 of 5 levels of health problems ((1) none, (2) slight, (3) moderate, (4) severe, or (5) extreme), with levels scored 1 through 5, respectively.

The EQ-5D-5L visual analog score is a self-assigned rating of overall health using a visual, vertical scale, with a score of 0 as the worst and 100 as best possible health. The EQ-5D-5L total

score and EQ-5D-5L visual analog score have been shown in many studies to be valid and reliable instruments for measuring HRQOL in subjects with GI diseases.

The scores for the 5 dimensions can be combined into a 5-digit number that describes the patient's health state. The health state index score will be calculated from the individual health profiles using country-specific value set.

The EQ-5D-5L will be conducted at Week -8 (Visit 1), Week 0 (Visit 3), Week 12 (Visit 4), Week 24 (Visit 6) or Early termination.

The change from baseline in EQ-5D-5L index score and EQ-VAS will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result at of last observed value before the first dose of study medication, i.e., the result at Week 0. The change from Week -8 to post-baseline values will also be analyzed, where Week -8 is defined as the first available result before treatment administration.

WPAI+CIQ:CeD

The WPAI+CIQ:CeD is a 10-question instrument questionnaire. The questionnaire assesses work time or academic classes lost due to CeD. Subjects also self-assess the impact of allergies on the performance in the workplace, at school, or during university classes. The recall period is the previous 7 days.

The WPAI+CIQ:CeD will be conducted at Week -8 (Visit 1), Week 0 (Visit 3), Week 12 (Visit 4), Week 24 (Visit 6) or early termination.

The derivation differs between subjects who are employed and subjects who are in school.

For those who are employed:

Percent work time missed due to CeD (Absenteeism) = Q2/(Q2+Q4)*100%

Percent impairment while working due to CeD (Presenteeism) = Q5/10

Percent overall work impairment due to CeD (Work Productivity Loss) = Q2/(Q2+Q4)+[(1-Q2/(Q2+Q4))x(Q5/10)]

Activity Impairment = Q10/10

The number of subjects employed, % of all subjects employed, number of work hours missed due to CeD and the parameters derived above will be summarized descriptively for those who are employed.

For those who are in school:

Percent work time missed due to CeD (Absenteeism) = Q7/(Q7+Q8)

Percent impairment while working due to CeD (Presenteeism) = Q9/10

Percent overall class impairment due to CeD (Class Productivity Loss) = Q7/(Q7+Q8) + [(1-Q7/(Q7+Q8))x(Q9/10)]

Percent activity impairment due to all health (Activity Impairment) = Q10/10
The number of subjects in school, % of all subjects in school, number of class hours missed due to CeD and the parameters derived above will be summarized descriptively for those who are in school.

The change from baseline in each item of WPAI+CIQ:CeD will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result at of last observed value before the first dose of study medication, i.e., the result at Week 0. The change from Week -8 to post-baseline values will also be analyzed, where Week -8 is defined as the first available result before treatment administration.

CeD-GSRS

The CeD-GSRS is a nondisease-specific 15-item instrument measuring the discomfort of GI symptoms using a modified Likert scale ((0) no discomfort at all, (1) minor discomfort, (2) mild discomfort, (3) moderate discomfort, (4) moderately severe discomfort, (5) severe discomfort, (6) very severe discomfort). In this study, the 10 items relevant to CeD will be evaluated (stomach pain/discomfort, hunger pains, nausea, rumbling in stomach, bloating, burping, passing gas, diarrhea, loose stools, urgent bowel movement).

The CeD-GSRS will be conducted at Week -8 (Visit 1), Week 0 (Visit 3), Week 12 (Visit 4), Week 24 (Visit 6) or early termination. The overall score is the mean of the 10 items. The smaller the total GSRS score, the milder the symptoms of the subject. If one item is missing, then the overall score will be set to missing.

The change from baseline in CeD-GSRS overall score will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result at of last observed value before the first dose of study medication, i.e., the result at Week 0. The change from Week -8 to post-baseline values will also be analyzed, where Week -8 is defined as the first available result before treatment administration.

CDAQ

The CDAQ is a validated, PRO measure developed to investigate the HRQOL of people living with CeD with a recall period of 4 weeks. This 32-item questionnaire measures 5 domains: stigma, dietary burden, symptoms, social isolation, and worries and concerns. Responses are measured using a 5-point Likert-type scale ((1) never, (2) rarely, (3) sometimes, (4) often, (5) always).

The CDAQ will be conducted at Week -8 (Visit 1), Week 0 (Visit 3), Week 12 (Visit 4), Week 24 (Visit 6) or early termination.

Table 6.f. CDAQ dimensions

Dimension	No. of questions	Question numbers
Stigma	8	2, 3, 4, 5, 6, 21, 23, 24
Dietary burden	8	25, 26, 27, 28, 29, 30, 31, 32
Symptoms	5	9, 10, 11, 12, 13
Social isolation	5	14, 16, 17, 18, 22
Worries and concerns	6	1, 7, 8, 15, 19, 20

The scores will be inverted following below table.

Table 6.g Original and inverted question scores

Response	Original score	Inverse score
Never	1	5
Rarely	2	4
Sometimes	3	3
Often	4	2
Always	5	1

For each dimension, the dimension score is calculated using the formula below and ranges from 0 to 100, where 0 indicates poorest quality of life and 100 indicates highest quality of life.

Stigma = (((Q2 + Q3 + Q4 + Q5 + Q6 + Q21 + Q23 + Q24) – 8) / 32) * 100

Dietary burden = (((Q25 + Q26 + Q27 + Q28 + Q29 + Q30 + Q31 + Q32) - 8) / 32) *100

Symptoms = (((Q9 + Q10 + Q11 + Q12 + Q13) – 5) / 20) * 100

Social isolation = (((Q14 + Q16 + Q17 + Q18 + Q22) – 5) / 20) * 100

Worries and concerns = (((Q1 + Q7 + Q8 + Q15 + Q19 + Q20) – 6) / 24) * 100

An overall index score will be calculated. As per the dimension scores, the overall index score ranges from 0 to 100, where 0 indicates poorest quality of life and 100 indicates highest quality of life. The formula for calculating the overall index score is:

Overall Index Score = (Stigma score + Dietary burden score + Symptoms score + Social isolation score + Worries and concerns score) / 5

If a respondent has not completed one or more questions, then the associated dimension score(s) and overall index score will not be calculated.

The change from baseline in CDAQ overall index score will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result at of last observed value before the first dose of study medication, i.e., the result at Week 0. The change from Week -8 to post-baseline values will also be analyzed, where Week -8 is defined as the first available result before treatment administration.

CD-QoL

The CD-QoL is a 20-item questionnaire measuring the quality of life in celiac patients over a 30-day time period and one overall quality of life item. Responses for item 1-20 are measured on a 5-point Likert-type scale ((1) not at all, (2) slightly, (3) moderately, (4) quite a bit, (5) a great deal). A lower score indicates lesser concern about QoL characteristic. Response for the overall questions is measured as (5) Excellent, (4) Very Good, (3) Good, (2) Fair, (1) Poor.

The CD-QoL will be conducted at Week -8 (Visit 1), Week 0 (Visit 3), Week 12 (Visit 4), Week 24 (Visit 6) or early termination.

The overall CD-QoL score will be calculated as the mean of the 20 items. A lower score indicates higher quality of life.

The change from baseline in CD-QoL score will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result at of last observed value before the first dose of study medication, i.e., the result at Week 0. The change from Week -8 to post-baseline values will also be analyzed, where Week -8 is defined as the first available result before treatment administration.

Serology titers:

Celiac Serology Titers

IgG DGP, IgA DGP, and IgA tTG results will be collected at Week -8 (Visit 1), Week 12 (Visit 5), Week 24 (Visit 6) and early termination.

The fold-change from baseline to Week 24 in celiac serology titers will be calculated as

fold-change from baseline = value at *Visit X* ÷ value at baseline

where baseline is defined as the last non-missing value prior to first dose of study drug.

Histological endpoints:

IEL

Tissue samples to evaluate IELs will be collected through duodenal biopsies at Week -4 (Visit 2) and at Week 24 (Visit 6) or early termination.

The change from baseline in IEL count will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result of last observed value before the first dose of study medication, i.e., the IEL count at Week -4.

Marsh-Oberhuber scores

Marsh-Oberhuber scores will be calculated from IEL and Vh:Cd results performed at Week -4 (Visit 2) and at Week 24 (Visit 6) or early termination. This will be done using the Q-MARSH scoring system described in Adelman et al. 2018, and modified according to experience at our Phase 2 vendor. Q-MARSH derives Marsh-Oberhuber scores of M0, M1, M2, M3a, M3b, and M3c from quantitative Vh:Cd and IEL count measurements. In addition, one final Marsh-Oberhuber score per biopsy reading will be given and reported as 0, 1, 2, 3a, 3b, and 3c.

As Marsh-Oberhuber is assessed and reported as a qualitative ordinal variable. The change from baseline in Marsh-Oberhuber scores, may be calculated as follows to estimate the mean number of Marsh-Oberhuber stage differences:

change from baseline = value at *Visit X* – value at baseline.

The change from baseline may also be calculated as the odds of changing from one stage to another using either a non-parametric estimate of the odds (or log odds) or a model based on an ordinal logistic regression model.

The baseline is defined as the result of last observed value before the first dose of study medication, i.e., the Marsh-Oberhuber scores at Week -4.

VCIEL

The VCIEL metric is a newly described calculation that takes the quantitative Vh:Cd ratio and IEL counts and combines them mathematically to generate a score that is continuous and not categorical[3]. The relevant equation is described in Syage et al (2023)[6] and is listed below. The VCIEL metric is expected to decrease in CeD patients treated with an efficacious therapy. This method provides a continuous quantitative metric for analysis.

$$VCIEL = \left[\frac{Vh{:}Cd}{\sigma_{Vh:Cd}} - \frac{IEL}{\sigma_{IEL}}\right]$$

Where $\sigma_{Vh:Cd}$ and σ_{IEL} are the standard deviations of Vh:Cd and IEL, respectively. In practice, the standard deviations (σ) are replaced by the sampled subjects' standard deviations, $s_{Vh:Cd}$ and s_{IEL}.

mpVCIEL

The multi-parametric VCIEL (mpVCIEL) metric is a vector-based measure of the L2-normalized Euclidean distance between two different distributions and takes the correlation between Vh:Cd and IEL into account. The change in mpVCIEL from baseline is defined as

$$\Delta mpVCIEL = \sqrt{[\Delta VhCd \quad \Delta IEL] * {S_{2-1}}^{-1} * \begin{bmatrix} \Delta VhCd \\ \Delta IEL \end{bmatrix}}$$

where $\Delta VhCd$ is the change from baseline for Vh:Cd and ΔIEL is the change from baseline for IEL and S_{2-1} is the sample estimate of the covariance of $\Delta VhCd$ and ΔIEL.

Urine GIP

During Weeks -2 through 24 or the end of study, whichever is earlier, subjects will collect their first morning void urine sample weekly for qualitative GIP. Collection must be done the day after consuming the SIGE bar or within 6-12 hours after ingesting a SIGE bar.

On the day of a clinic visit, subjects will collect the first morning void in CHAPS-coated cups and take the sample to the site for quantitative GIP. Collection must be done the day after the SIGE bar consumption or within 6-12 hours after ingestion of a SIGE bar.

The change from baseline in quantitative GIP will be calculated as

change from baseline = value at *Visit X* – value at baseline,

where baseline is defined as the result of last observed value before the first dose of study medication, i.e., the quantitative GIP at Week 0.

6.5.3.2 Main Analytical Approach

The endpoint of change from baseline in CDSD symptom severity item score for each individual symptom and change in each of the frequency questions (bowel movements, diarrhea, and vomiting) from Week -1 and from Week -8 to Weeks 12 and 24 will be analyzed using the same approach as the primary analysis described in Section 6.5.1.2. Week -8 is defined as the result of the first week before treatment administration. *The frequency of bowel movements will be converted to an indicator for constipation (defined as fewer than 3 days of complete spontaneous bowel movement in a week). The weekly average of the frequency items will be an average of each daily frequency during the week, with constipation having values of 0 (no constipation) or 1 (constipation) each week. Weekly frequency will be calculated where the CDSD is completed on at least 4 of the 7 days. When the daily diary is completed on less than 4 of the 7 days, the weekly frequency will be considered missing. The change in symptom severity and frequency for each item in CDSD from Week -3 (before SIGE) and from Week -5 (before study intervention) to*

Weeks 12 and 24 will also be analyzed. Week -5 will be defined as the week prior to run-in start date. Week -3 will be defined as the second week of run-in period. The analysis method will be the same as for the corresponding change from Week -1 and change from Week -8. Nominal p-values will be calculated. No hypothesis testing will be performed.

PGIC

Raw and collapsed PGIC score and dichotomized (Improved vs Not Improved) at Weeks 12 and 24 and the change from baseline in PGIS score will be assessed using ANCOVA (proportional odds model for PGIC dichotomized variable) with treatment group and the randomization stratification factors as covariates.

CDSD and other PRO endpoints

For each subject, the change from baseline in CDSD score of subject's most bothersome symptom at baseline will be calculated. The CDSD score of the most bothersome symptom will be analyzed using the same approach as Section 6.5.1.2.

In addition, the change from baseline in CDSD score of subjects who reported their most bothersome symptom as diarrhea at baseline will be summarized separately. Same analysis will be conducted for nausea, abdominal pain and other most bothersome symptoms at baseline.

Other exploratory PRO endpoints at Weeks 12 and 24 will be analyzed using the same approach as in Section 6.5.1.2. This includes ICDSQ domain/overall score, PROMIS-Cognitive Function Instrument t-score, PROMIS-Fatigue Instrument t-score, SF-12 v2 domain/summary score, EQ-5D-5L index score, EQ-VAS, WPAI+CIQ:CeD, CeD-GSRS overall score, CDAQ overall index score and CD-QoL score.

Celiac Serology Titers

The fold change from Week -4 to Week 12 and Week 24 in celiac serology titers will be summarized by Visit and treatment group and analyzed using the same approach as in Section 6.5.1.2.

Marsh-Oberhuber and Q-MARSH scores

Marsh-Oberhuber scores are ordinal qualitative scores and will be summarized and assessed using a cumulative logit model with a proportional odds structure with treatment group, baseline result and the randomization stratification factors as covariates. A means model may also be used to investigate the mean number of stages changed from baseline.

While Q-MARSH scores are quantitative, they cannot be considered as interval variables and score change from baseline will be evaluated using the models described for Marsh-Oberhuber.

Histological endpoints, PRO and histological endpoints

Other exploratory histology efficacy endpoints at Week 24, including change from baseline in histology endpoints such as Vh, Cd, IEL, mpVCIEL, quantitative Marsh, VCIEL, [redacted] *for the first tertile of the intestine will be analyzed using the same approach as in Section 6.5.2.2.*

For binary endpoints, number and proportion of subjects will be provided for each treatment arm. The comparison between each dose of TAK-062 with placebo will be performed using the Cochran-Mantel-Haenszel test adjusted by the randomization stratification factors. The point p-values will be provided. Non-responder imputation will be used for missing data.

All efficacy endpoints will also be summarized by descriptive statistics by treatment group at each visit.

[redacted]

GIP endpoints

The frequency of positive urine GIP will be summarized by treatment arms for the second half of the run-in period (Week -2 to Day -1) and every week during the treatment period. The change in the frequency of positive urine GIP from the run-in period to each of the two-week durations in the treatment period will be summarized by treatment arms.

The frequency versus time and the change in frequency versus time will be plotted by treatment arms.

The proportion of subjects with positive urine GIP will be summarized by treatment arms for each visit from Visit 1 (screening) to Visit 6 (Week 24) as well as early termination visit.

6.5.3.3 Supplementary Analysis

Supplementary analyses based on subjects who satisfied study entry criteria and are on pre-enrollment diet will be performed. The corresponding endpoints will be analyzed based on FAS-no SIGE.

6.5.4 Subgroup Analyses

The change from baseline (Week -8 and Week -1) to Week 12 and Week 24 in CDSD GI symptom severity score will be summarized descriptively by treatment arm for each subgroup

defined in Table 6.h if applicable. A forest plot will be generated to display the result of each analysis. When there are at least 10 subjects in each treatment arm in a subgroup, primary efficacy analysis will be performed on the subgroup. Missing data will not be imputed.

Table 6.h Subgroups of Interest

Subgroup of Interest	Subgroup Categories
Age	12-<18, 18-<40, 40 or older Adolescent vs adult
Sex	Female vs Male
Race	Non-Hispanic White vs Hispanic / Non-white
Baseline celiac serology status	Seronegative (IgA tTG ≤ ULN and IgG DGP ≤ 2xULN) vs seropositive (IgA tTG > ULN or IgG DGP > 2xULN)
Baseline Symptom severity	Mild – Moderate (Vh :Cd ≥ 1.5) vs Moderate – Severe (Vh :Cd < 1.5)
PPI/H2RA use at Consent	PPI/H2RA vs no PPI/H2RA
Baseline symptom type	Diarrhea vs. no diarrhea
Baseline severity of small intestine mucosal injury	Vh:Cd < mean (Vh :Cd) vs >= mean (Vh :Cd)
Gluten exposure by GIP urine evaluation during the run-in period	>1 positive sample vs 0-1 positive sample

6.6 Safety Analysis

All safety analyses will be conducted using the SAF, the SAF-SIGE, and the SAF–no SIGE according to the actual treatment received. No statistical testing or inferential statistics will be generated. All AEs will be coded using the latest MedDRA version. The number and percentage of subjects with TEAEs (new onset or worsening AEs after the first dose of study treatment regardless of relationship to study drug), treatment-related AEs, SAEs, and AEs leading to treatment discontinuation will be summarized by MedDRA primary System Organ Class and Preferred Term, overall and by severity for each treatment group.

Change from baseline in clinical laboratory tests and vital signs will be summarized by treatment group and by visit. Subjects with markedly abnormal values for laboratory tests and vital signs will be summarized and listed.

6.6.1 Adverse Events

All AEs will be coded using the latest version of MedDRA. All AEs 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 an end date missing or

on or after the first dose will be summarized as TEAEs regardless of severity and relationship to study medication.

The following summaries will be presented:

- Overview of TEAEs during the study - number and percentage of subjects, number of events.
- TEAEs by system organ class (SOC) and preferred term (PT) - number and percentage of subjects.
- Severity of TEAEs by SOC and PT - number and percentage of subjects.
- Relationship of TEAEs by SOC and PT - number and percentage of subjects.
- TEAEs leading to study discontinuation by SOC and PT - number and percentage of subjects.
- Serious TEAEs by SOC and PT - number and percentage of subjects.
- Serious TEAEs by relationship to study drug and by SOC and PT - number and percentage of subjects.
- Serious TEAEs by intensity and by SOC and PT - number and percentage of subjects.
- Most frequent TEAEs (>=5% of subjects in any treatment group, sorted by frequency) by PT - number and percentage of subjects.
- TEAEs resulting in death by PT - number and percentage of subjects.

Key guidelines for determining the incidence of AEs are as follows:

- AEs with missing or unknown intensity will be considered as severe.
- AEs with missing or unknown relationship to study drug will be counted as related.
- A subject with 2 or more AEs within the same level of the MedDRA term will be counted only once in that level.
- SOCs will be sorted in alphabetical order. 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.
- For the summary of TEAEs by SOC and PT and intensity, if a subject experiences more than 1 episode of a particular coded AE, 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 toxicity grade in that SOC.
- In selected summaries, adverse events will be summarized by the number of events reported in addition to the number and percentage of subjects with events.

Data listings for TEAEs, TEAEs leading to study discontinuation, SAEs and deaths will be presented. The AEs 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, action taken concerning study drug, causality to study drug, the outcome, whether the adverse event was an SAE.

6.6.2 Adverse Events of Special Interest

There are no AEs of special interest considered for TAK-062.

6.6.3 Clinical Laboratory Evaluations

Blood samples for analysis of the Hematology, Serum Chemistry, Serum and Urine parameters will be summarized by visit and by treatment group using the SAF, the SAF-SIGE, and the SAF–no SIGE.

- Summary statistics (n, mean, SD, median, minimum, and maximum) by treatment group and overall for the actual values and change from baseline values.
- Markedly abnormal values (MAVs), as defined in Section 9.3, 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, shift tables for certain clinical laboratory parameters of interest from baseline to worst post-baseline value will be presented. Each subject will be categorized as described in Table 6.i for the baseline value and worst post-baseline value. The number of subjects in each of the combinations of shifts will be presented.
- For female subjects, pregnancy test results will be listed by treatment, study center and subject number.

Table 6.i Categories of Worst Post-baseline Values

Parameter	Category 1	Category 2	Category 3	Category 4	Category 5
ALT	>ULN and ≤3 × ULN	>3 × ULN and ≤5 × ULN	>5 × ULN and ≤8 × ULN	>8 × ULN and ≤20 × ULN	>20 × ULN
AST	>ULN and ≤3 × ULN	>3 × ULN and ≤ 5 × ULN	>5 × ULN and ≤8 × ULN	>8 × ULN and ≤20 × ULN	>20 × ULN
ALT or AST (either meeting the criteria)	>ULN and ≤3 × ULN	>3 × ULN and ≤ 5 × ULN	>5 × ULN and ≤8 × ULN	>8 × ULN and ≤20 × ULN	>20 × ULN

Parameter	Category 1	Category 2	Category 3	Category 4	Category 5
ALT/AST and Total Bilirubin	ALT>ULN and ≤3 × ULN and AST>ULN and ≤3 × ULN and TBILI>ULN and ≤2 × ULN	ALT>ULN and ≤3 × ULN and AST>ULN and ≤3 × ULN and TBILI>2 × ULN	(ALT>3 × ULN or AST>3 × ULN) and TBILI>2 × ULN		

* ULN: Upper Limit of Normal

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 Système International (SI) units.

6.6.4 Vital Signs

Vital signs and weight at scheduled visits and their changes from baseline will be summarized by visit for each treatment group and overall using descriptive statistics. Subjects with MAVs for vital sign values, as defined in Section 9.3, will be tabulated.

A listing of all vital signs and weight data will be provided. Values meeting the MAV criteria will be flagged.

6.6.5 12-Lead ECGs

Electrocardiogram (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 MAV ECG value, as defined in Section 9.3, will be tabulated for each variable across all visits.

A listing of all ECG data will be provided. Values meeting the MAV criteria will be flagged.

6.6.6 Immunogenicity

Each ADA sample will be tested first by a screening test, followed by a confirmatory test if the screening test is positive. An ADA test will be considered positive only when both the screening and confirmatory tests show positive results. Testing for neutralizing antibodies (NAbs) will be performed only if the ADA testing is considered positive. A positive ADA subject is defined as a subject who has at least 1 positive ADA result during the study, and is further categorized as:

- *Transiently positive: defined as subjects with confirmed positive ADA in at least 1 sample and no consecutive samples.*
- *Persistently positive: defined as subjects with confirmed positive ADA in 2 or more consecutive positive ADA samples.*

Immunogenicity will be analyzed using the immunogenicity analysis set. Percentage of subjects with positive ADA blood sample results and percentage of subjects with positive neutralizing ADA during the study will be summarized by treatment group. Percentage of subjects with positive ADA will be also summarized by study visit and treatment group. The impact of positive ADA on PK, efficacy, and safety may be evaluated. Results of ADA and NAb tests will be also in listings.

6.6.7 Extent of Exposure and Compliance

The SAF, the SAF-SIGE, and the SAF-no SIGE will be used for all summaries in this section. Study drug/ SIGE bar exposure, compliance for study drug, and compliance for SIGE bar will be summarized by visit and by treatment group.

- The extent of exposure will be calculated as the duration between the first and last dose of study drug as following:

  Date of last dose of study treatment – Date of first dose of study treatment + 1

- Compliance for drug/SIGE bar will be calculated as the following:
  - Number of tablets taken divided by the expected number of tablets to be taken
  - Number of SIGE bar taken out of the total number of expected SIGE bars

Study drug administration data including placebo, TAK-062, and SIGE bar will be presented in data listings.

6.7 Pharmacokinetic, Pharmacodynamic, and Biomarker Analyses

6.7.1 Pharmacokinetic Analysis

Measured plasma concentrations of TAK-062 over time will be summarized descriptively using the PK analysis set. Subjects in PK analysis set will be analyzed according to the treatment regimen they received and by study visit. *Individual concentration data versus time will be presented in a data listing.*

Further analysis may be performed as deemed necessary and will be reported separately from the clinical study report (CSR).

PK parameters will be summarized for each cohort using descriptive statistics (n, mean, SD, %CV, geometric mean, geometric mean %CV, median, minimum, maximum) by nominal dose group, and visit if applicable.

Other PK parameters, as appropriate may be determined or calculated at the discretion of the independent clinical pharmacologist. An ad-hoc PK analysis may be conducted in blinded manner after Cohort 1.

6.7.2 Biomarker Analysis

The biomarker analysis will be detailed in a separate analysis plan and reported in a separate document.

6.8 Patient Reported Outcomes (PROs) and Health Care Utilization Endpoints Analysis

6.8.1 PRO Analysis

Analyses of PRO endpoints are described as part of the efficacy analysis in Section 6.5.

Additional analyses for PRO validation will be analyzed separately. A separate report will be created.

6.8.2 Health Care Utilization Analysis

Not applicable.

6.9 Other Analyses

Not applicable.

6.10 Interim Analyses

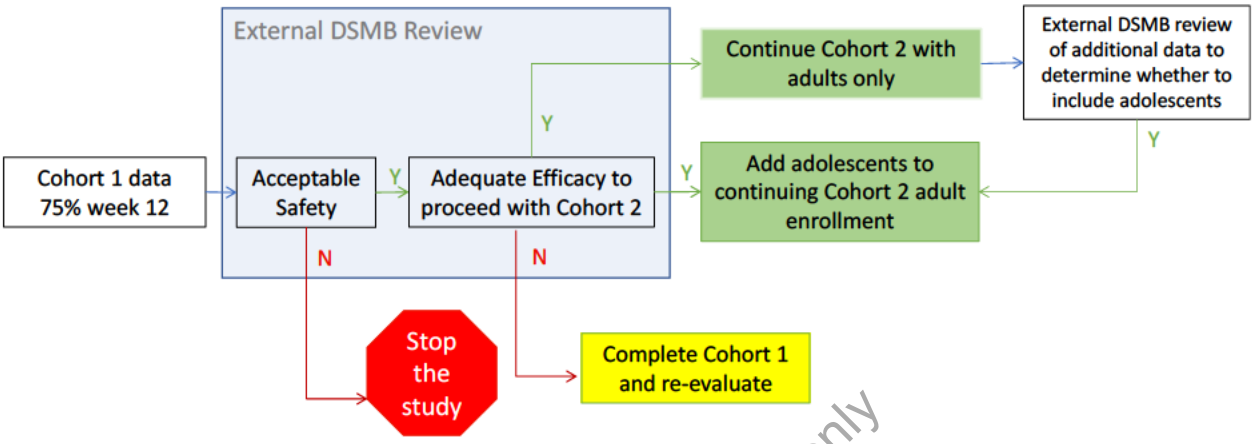
An IA is to be performed to assess if clinical efficacy and safety have been demonstrated. The IA will be performed when at least 75% of subjects in Cohort 1 have either completed 12 weeks of treatment or dropped out. The IA will include safety analysis, primary efficacy analysis, and histological data analysis based on available data. The IA will be used to decide on the addition of adolescents in Cohort 2. The decision will be based on the safety evaluation, the conditional power (CP) for the primary efficacy analysis and the analysis of Vh:Cd data. There is no plan to stop the study early for efficacy or futility based on the interim efficacy analysis results. Thus, there will be no alpha level adjustment for the final analysis.

The IA will be performed by an external independent data monitoring committee that maintains the blind of the sponsor TAK-062 study team, investigators, site staff, and study subjects to the study treatment assignment until the database has been locked and unblinded. Specific procedures and guidelines on blinding and unblinding and controlled dissemination of interim analysis results are described in the Data Access Management Plan and/or IDMC charter.

The study will be completed as planned unless 1 or more of the following criteria are satisfied that require temporary suspension or early termination of the study.

- New information or other evaluation regarding the safety or efficacy of TAK-062 that indicates a change in the currently known benefit-risk profile such that the benefit-risk balance is no longer acceptable for subjects participating in the study.
- Significant violation of GCP that compromises the ability to achieve the primary study objectives or compromises subject safety.

Following decision tree for the IA will be followed.



DSMB: Data safety monitoring board

The safety analysis at IA will use SAF population as of data cutoff date of IA; the efficacy analysis will use FAS-SIGE population as of data cutoff date of IA.

Safety Considerations

Safety criteria will focus on the severity of reported adverse events. When the proportion of subjects with grade >=3 related events is more than 20% (percentage point) higher in the active arm than in the placebo arm, the safety will be considered not acceptable and may lead to early termination of the study.

Efficacy Considerations

For efficacy, the recommendation chart with conditional powers for CDSD GI symptom severity score and the probability of Vh:Cd will be used for decision making in conjunction with the clinical evaluation of the totality of the study data.

The CP for CDSD GI symptom severity score will be calculated based on the test statistic in the MMRM model of the primary efficacy endpoint at Week 12 on the available CDSD observations up to Week 12. Conditional power is the probability of achieving statistical significance (2-sided type-I error rate of 0.05) at the final analysis given data observed at interim analysis and assuming that future data are generated under a true effect of observed or a hypothesized effect size. In this study, the CP will be calculated assuming that future data will be generated under the observed effect size and is calculated as

$$CP = \Phi\left(\frac{\frac{b}{\sqrt{I}} - Z_{(1-\alpha/2)}}{\sqrt{1-I}}\right)$$

where

I = information fraction at time of IA and is calculated as

$$I = \frac{\frac{1}{N_1} + \frac{1}{N_2}}{\frac{1}{n_1} + \frac{1}{n_2}}$$

$b = Z_I \cdot \sqrt{I}$ and Z_I is the observed MMRM test statistic at interim analysis

n_1 = sample size in placebo arm at IA

n_2 = sample size in TAK-062 [redacted] arm at IA

N_1 = total sample size in placebo arm at final analysis

N_2 = total sample size in TAK-062 [redacted] arm at final analysis

$Z_{(1-\alpha/2)} = \left(1 - \alpha/2\right)^{th}$ quantile of the standard normal distribution

$\Phi(\cdot)$ = cumulative probability of the standard normal distribution

The level of conditional probability of Vh:Cd is determined using ANCOVA as specified for the secondary endpoint in this SAP based on the available Vh:Cd observations at Week 24 from the IA data snapshot. The conditional probability level is determined by the lower limit of 1- sided confidence internal (CI) for change from baseline of VhCd from the ANCOVA results using the general linear model.

P(effect size $\geqslant$ 0.1) $\geqslant$ 75%	Lower Limit (LL) of 1-sided 75% CI of effect size $\geqslant$ 0.1
P(effect size $\geqslant$ 0.1) $\geqslant$ 60%	LL of 1-sided 60% CI of effect size $\geqslant$ 0.1
P(effect size $\geqslant$ 0) $\geqslant$ 50%	LL of 1-sided 50% CI of effect size $\geqslant$ 0
P(Tx CFB $\geqslant$ 0.1) $\geqslant$ 50%	LL of 1-sided 50% CI in Tx arm $\geqslant$ 0.1

The recommended decision criteria based on the CP for CDSD GI symptom severity score and probability of Vh:Cd, key endpoints, after safety assessment, is shown below, however the DMC is empowered to make a recommendation based on the totality of data.

The updates for efficacy considerations were incorporated into the DMC charter prior to the interim analysis and reflected in SAP version 3 to ensure consistency.

Conditional Power and Efficacy Decision Matrix

		Symptom Efficacy (difference between arms in CFB of CDSD GI severity score at Week 12)				
		CP[a] ⩾75%	**60%⩽ CP <75%**	**30%⩽ CP <60%**	**10%⩽ CP <30%**	**CP <10%**
Histology efficacy (difference between arms in CFB of Vh:Cd[b] at Week 24)	P (effect size ⩾ 0.1) ⩾75%	High Probable Treatment Effect	Intermediate Probable Treatment Effect	Intermediate Probable Treatment Effect	Intermediate Probable Treatment Effect	Intermediate Probable Treatment Effect
	P (effect size ⩾ 0.1) ⩾60%	High Probable Treatment Effect	Intermediate Probable Treatment Effect	Intermediate Probable Treatment Effect	Intermediate Probable Treatment Effect	Low Probable Treatment Effect
	P (effect size ⩾ 0) ⩾50%	Intermediate Probable Treatment Effect	Intermediate Probable Treatment Effect	Intermediate Probable Treatment Effect	Low Probable Treatment Effect	Low Probable Treatment Effect
	P (Tx CFB ⩾ 0.1) ⩾ 50%	Intermediate Probable Treatment Effect	Intermediate Probable Treatment Effect	Intermediate Probable Treatment Effect	Low Probable Treatment Effect	Low Probable Treatment Effect
	None of the above	Intermediate Probable Treatment Effect	Low Probable Treatment Effect	Low Probable Treatment Effect	Low Probable Treatment Effect	Low Probable Treatment Effect

CDSD: Celiac Disease Symptom Diary; CFB: change from baseline; CP: conditional probability; P: probability of Vh:Cd; Tx: active treatment (TAK-062) group; Vh:Cd: villous height to crypt depth ratio

[a] CP: the probability that the final primary efficacy analysis based on 12-week CDSD will be statistically significant given the observed CDSD.

[b] Vh:Cd: Villus height to crypt depth ration.

Note: 'Effect size' in Histology Efficacy decision criteria is defined as the treatment difference in change from baseline of Vh:Cd at week 24.

The decision criteria will be used, in conjunction with the DMC review of totality of data, to determine whether to proceed with enrollment of adolescents based on the CP for CDSD GI symptom severity score and conditional probability of Vh:Cd, after ruling out significant safety concerns.

If interim analysis results indicate probable treatment effect in the 'High' (probable treatment effect) or 'Intermediate' (probable treatment effect) treatment effect zones, and there are no safety concerns, the study may proceed with enrollment of cohort 2, per protocol and cohort 2 may include enrollment of adolescents. If interim analysis results indicate probable treatment effect in the 'Low', and there are no safety concerns, the study may proceed with completion of cohort 1 (i.e., last subject in cohort 1 completes visit 7 [week 28]) and will reassess full cohort 1 data before initiating cohort 2 enrollment.

6.11 Independent Data Monitoring Committee

According to ICH E6 (1.25) an independent data monitoring committee (eg, Independent Data Monitoring Committee, Data and Safety Monitoring Committee, Monitoring Committee, Data Monitoring Committee) may be established by the sponsor to assess at intervals the progress of the study, the safety data, and the critical efficacy endpoints, and to recommend to the sponsor whether to continue, modify, or stop a study.

The minimal description of the independent data monitoring committee should include:

- Composition (number of members and expertise).
- Frequency of meetings/assessments.
- Purpose and authority.
- Scope of data review.

Details of the Independent Data Monitoring Committee (IDMC) will be captured in a charter prior to the start of the trial.

An external IDMC is warranted given the inclusion of adolescents and the current stage of development.

The primary responsibility of the IDMC is to safeguard study subjects, in particular the adolescent subjects, by reviewing and assessing the clinical safety data being collected during the performance of the study. The IDMC will review SAEs and AEs ⩾Grade 3. The IDMC will also meet periodically throughout the study to review cumulative safety data. Based on these data evaluations, the IDMC will make recommendations to the sponsor to continue the study as planned, or to modify, temporarily suspend, or terminate the treatment group or study. The IDMC will be responsible for identifying issues and making recommendations regarding the monitoring of the subjects for safety, including the collection of additional safety data. IDMC will also be reviewing IA data and notifying the sponsor the recommendation to include adolescent group in Cohort 2. The sponsor will be responsible for notifying investigators and regulatory authorities of any IDMC recommendations, as appropriate.

7.0 REFERENCES

2019a. PROMIS Cognitive Function Scoring Manual.
2019b. PROMIS Fatigue Scoring Manual.
Coeliac Disease Assessment Questionnaire (CDAC) Scoring Guide 2016
Adelman, D. C., Murray, J., Wu, T. T., Maki, M., Green, P. H. and Kelly, C. P. 2018. Measuring change in small intestinal histology in patients with celiac disease. Am J Gastroenterol, 113(3), 339-47
Syage, J.A., et al., A Composite Morphometric Duodenal Biopsy Mucosal Scale for Celiac Disease Encompassing both Morphology and Inflammation. Clinical Gastroenterology and Hepatology, 2023
Taavela, J., Koskinen, O., Huhtala, H., Lahdeaho, M. L., Popp, A., Laurila, K., et al. 2013. Validation of morphometric analyses of small-intestinal biopsy readouts in celiac disease. PLoS One, 8(10), e76163.

8.0 CHANGES TO PROTOCOL PLANNED ANALYSES

None.

9.0 APPENDIX

9.1 Changes From the Previous Version of the SAP

SAP v3.0 Summary of Changes Since the Last Version of the Approved SAP			
Change Number	**Sections Affected by Change**	**Description of Each Change and Rationale**	
	Location	***Description***	***Rationale***
1.	Section 1.2.4, Section 6.5.3	Add more Exploratory/Additional Endpoints	[redacted]
2.	Section 1.3	Strategy for Intercurrent event was updated	[redacted]
3.	Section 6.5.1.4	Added additional sensitivity analysis	[redacted]
4.	Section 6.10	Decision criteria were updated and details were added for determining the conditional probability for Vh:Cd	Clarification

9.2 Data Handling Conventions

9.2.1 General Data Reporting Conventions

Means and medians will be presented to 1 more decimal place than the recorded data. The standard deviations (SDs) will be presented to 2 more decimal places than the recorded data. CIs intervals will be presented using the same number of decimal places as the parameter estimate.

9.2.2 Definition of Baseline

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, unless stated otherwise. 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].

In general, baseline is defined as the last non-missing measurement prior to the first dose of study drug or, for participants who are randomized but not dosed, the randomization date, unless stated otherwise.

9.2.3 Definition of Visit Windows

Subjects do not always adhere strictly to the visit timing stated in the protocol. Therefore, the designation of visits will be based on the day of evaluation relative to the start of study drug rather than the nominal visit recorded in the data. Accordingly, the study is divided into continuous, mutually exclusive analysis windows. Unless otherwise stated, all analysis will be based on the analysis visit.

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 9.a, Table 9.b, and Table 9.c.

If a subject has more than one measurement included within a window, the assessment closest to the target day will be used. In case of ties between observations located on different sides of the target day, the later assessment will be used. In case of ties located on the same side of the target day (i.e., more than one value for the same day), the mean of the values will be used.

9.2.3.1 Study Week/Visit Windows for Efficacy Data

The visit windows for the analysis of efficacy data are defined in Table 9.a.

Table 9.a Visit windows for the analysis of Primary endpoint (CDSD GI symptom severity score)

Visit	Target Day	Day Range
Week -8 (Visit 1)	**Day -56**	first week upon study entry
Week -5	**Day -35**	the week before the run-in date
Run-in start date/Week -4 (Visit 2)	**Day -28**	min(first placebo IP date in dosing diary, max(EGD date, V2 date))
Weeks 1 – 24 *		Every 7 days from treatment start date (Day 1)

*Analysis week could differ from analysis day. To derive analysis week, first derive analysis day 1 and run-in start date, then derive analysis week based on these days.

Table 9.b Visit windows for the Analysis of PRO data

Visit	Target Day	Day Range	
		PGIS	**Other PRO questionnaires [a]**
Week -8 (Visit 1)	**Day -56**	≤ -43	≤ -43
Week -4 (Visit 2)	**Day -28**	-42 – -3	-42 – -3
Week 0 (Visit 3)	**Day 1**	-2 – 1	-2 – 1
Week 3 (Visit 4)	**Day 21**	2 – 40	
Week 12 (Visit 5)	**Day 84**	41 – 127	2 – 127
Week 24 (Visit 6)	**Day 168**	128 – 182	128 – 182
Week 28 (Visit 7)	**Day 196**	≥183	≥183

[a] The windows will be applied to following questionnaires: EQ-5D-5L, CeD Most Bothersome Symptom, PROMIS-Fatigue Instrument, CeD-GSRS, PROMIS–Cognitive Function Instrument, WPAI:SHP–CeD, SF-12, ICDSQ, CDAQ, CD-QOL, PGIC.

Note: Subjects terminated early, their assessment (if recorded as ET) may be assigned to appropriate weeks according to above windows.

9.2.3.2 Visit Windows for the Analysis of Safety Data

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 24 value may be different than the End of Treatment Visit value.

Table 9.c Visit Windows for the Analysis of Safety Data

Visit	Target Day	Day Range		
		Vital Sign, ECG, EGD	Hematology, Chemistry, Urinalysis, hCG, Urine sample for GIP	Celiac serologies, [redacted]
Baseline	**Day 1**	≤1	≤ 1	≤ 1
Week 3 (Visit 4)	**Day 21**		2 – 40	
Week 12 (Visit 5)	**Day 84**		41 – 126	2 – 126
Week 24 (Visit 6)	**Day 168**	≥2	127 – 182	≥127
Week 28 (Visit 7)	**Day 196**		≥183	

(a) For some assessments, baseline is collected on Study Day 1 before study drug administration.
Note: Subjects terminated early, their assessment (if recorded as ET) will be assigned to appropriate weeks according to above windows.

9.2.4 Conventions for Missing Adverse Event Dates

Every effort will be made to determine the actual onset date for the event or to obtain a reliable estimate for the onset date from the investigator.

For AEs or SAEs, a missing or incomplete onset date will be imputed according to the following conventions:

1. If an onset date is missing, the derived onset date will be calculated as the first non-missing valid date from the following list (in order of precedence):
   - First study medication date.
   - Consent date (for SAEs only).
2. If an onset date is incomplete, the derived onset date will be calculated according to the following:
   - Missing day, but month and year present: the day will be imputed as the 1st of the month. If the first study drug dose occurs in the same month and year but after the imputed date, the derived onset date will be set equal to the first study drug date.
   - Missing day and month, but year present: the day and month will be imputed as the 30th June of the year. If the first study drug dose occurs in the same year but after the imputed date, the derived onset date will be set equal to the first study drug date. If

the EOS date occurs in the same year but before the imputed date, the derived onset date will be set equal to the first study drug date.

For AEs or SAEs, a missing or incomplete end date will be imputed according to the following conventions:

1. If an end date is missing, the derived end date will be imputed as the last assessment date, assuming that the last assessment occurs after the AE start. If the last assessment occurs prior to the AE start date, the derived end date will be imputed as the AE start date.
2. If an end date is incomplete, the derived end date will be calculated according to the following:
   - Missing day, but month and year present: the day will be imputed as the last date (for example February 2009 will be imputed as 28 February 2009) of the month.
   - Missing day and month, but year present: the day and month will be imputed as the 31st December of the year.
   - If the imputed AE end date occurs after the database lock date, the imputed AE end date will be imputed as the database lock date.

9.2.5 Conventions for Missing Concomitant Medication Dates

Start and stop dates for medication history and concomitant medications are collected on the eCRF. Definitions of medication history and concomitant medications are defined in Section 6.4. In case of missing or partial dates for concomitant medications, or medication history, the following rules will be used:

If the start date is partial or unknown:

1. If the day is missing, the start day will be the first day of the month.
2. If the month is missing:
   - If the year is the same as the year of first dose of study drug, the start month will be the month corresponding to 90 days prior to the date of first dose of study drug with exception that the month of first dose is Jan, Feb, or Mar.
   - If the year is the same as the year of first dose of study drug and the month of first dose is Jan, Feb, or Mar, the start month will be Jan.
   - If the year is not the same as the year of first dose of study drug, the start month will be Jan.
3. If the entire date is unknown (e.g., the year is missing):
   - If eCRF indicates that the medication ended prior to the informed consent date, then the medication start date will be imputed to the informed consent date minus one day.
   - Otherwise, the start date will be minimum of the date of first dose of study drug and the medication end date.

If the stop date is partial, unknown or "ongoing":

1. If the day is missing, the stop day will be the last day of the month reported.
2. If the month is missing:
   - If the year is the same as the year of last assessment, then the stop month will be to the month during which the last assessment occurred.
   - If the year is not the same as the year of the last assessment, then the end month will be Dec.
3. If the entire date is unknown (e.g. the year is missing) or if the medication is "ongoing", the stop year will be the year in which the last assessment occurred. If information collected on the eCRF indicates that the medication ended prior to the informed consent date, then the medication stop date will be imputed as the informed consent date minus one day.

If both start date and end date are imputed and the imputed start date is greater than the imputed end date, then the imputed start date will be set to the imputed end date.

No dates will be imputed for previous medications.

9.3 Criteria for Markedly Abnormal Values

Table 9.d Criteria for Markedly Abnormal Values for Laboratory Parameters

– **Parameter**	– **Unit**	–	– **Lower Criteria**	– **Upper Criteria**
– Prolactin	–	–	– -	– >5 x ULN
– Albumin	– g/L	–	– <20	– -
– Total protein	–	–	– <0.8 x LLN	– >1.2 x ULN
– Creatinine	–	–	– -	– >1.5 x ULN
– Blood urea nitrogen	–	–	– -	– >3 x ULN
– Creatine kinase (CPK)	–	–	– -	– >5 x ULN
– Potassium	– mmol/L	–	– <3.0	– >5.5
– Sodium	– mmol/L	–	– <130	– >155
– Calcium	– mmol/L	–	– <2.0	– >2.9
– Chloride	–	–	– <0.8 x LLN	– >1.2 x ULN
– Magnesium	– mmol/L	–	– <0.4	– >1.23
– Phosphate	– mmol/L	–	– <0.6	–
– Total cholesterol	– mmol/L	–	– -	– >10.34
– Triglycerides	– mmol/L	–	– -	– >5.7
– High-density lipoprotein cholesterol	–	–	– <LLN	– -
– Low-density lipoprotein cholesterol	–	–	– -	– >ULN
– Glucose	– mmol/L	–	– <2.2	– >13.9
– 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	– x 10^9/L	–	– <2.0	– >100
– Neutrophils	– x 10^9/L	–	– <1.0	– -
– Lymphocytes	– x 10^9/L	–	– < 0.5	– > 20

– Hemoglobin	– g/L	–	– <80	– >40 above ULN
– Hematocrit	–	–	– <0.8 x LLN	– >1.2 x ULN
– Platelets	– x 10^9/L	–	– <75.0	– >600
– Mean corpuscular hemoglobin	–	–	– <0.8 x LLN	– >1.2 x ULN
– PT	–	–	– -	– >2.5 x ULN
– INR	–	–	– -	– >2.5 x ULN

Table 9.e Criteria for Markedly Abnormal Values for Vital Signs Parameters

– **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	– >95 or increase from baseline of >10
– Body temperature	– ºC	– <35.0	– >38.3
–	–		–

Table 9.f Criteria for Markedly Abnormal Values for 12-Lead ECG Parameters

– **Parameter**	– **Unit**	– **Lower Criteria**	– **Upper Criteria**
– Heart rate	– bpm	– <50 or decrease from baseline of >20	– >120 or increase from baseline of >20
– PR	– msec	– <120	– >200
– QTcF Interval	– msec	– -	– [>500] **or** – [>30 increase from baseline and >450]
– QRS	– msec	– <60	– >120

9.4 Analysis Software and Sample codes

The below software will be used in the analyses:

SAS® Version 9.4, or higher.

MMRM Sample codes:

```
proc mixed  /* select derived data set (all related weeks) */ ;
class trt subjid week strata;
model chg = trt week strata trt*week baseline/ddfm=kenwardroger;
repeated week / subject=subjid type=UN;
random intercept / subject=subjid type=un;
lsmeans trt trt*week /cl diff ;
run;
```

ANCOVA Sample codes:

```
proc glm /* select derived data set  */ ;
class trt strata;
model chg = trt strata base/alpha=.05;
lsmeans trt/stderr pdiff diff;
run;
```

Multiple Imputation Sample codes:

1. For primary endpoint

```
proc mi data=cdsdwk_t out=imputed seed=128 nimpute=1000
 minimum= . . . . 0 0 0 0 0 0 0 0 0 0 0 0 0
 maximum= . . . . 4 4 4 4 4 4 4 4 4 4 4 4 4/* double check the range 0-4*/;
 class trt strat1 strat2 strat3;
 var  strat1 strat2 strat3 week0 week1 week2 week3 week4 week5 week6 week7 week8 week9
week10 week11 week12;
fcs reg (week0= strat1 strat2 strat3);
 fcs reg (week1= strat1 strat2 strat3 week0);
 fcs reg (week2= strat1 strat2 strat3 week0 week1);
 fcs reg (week3= strat1 strat2 strat3 week0 week1 week2);
 fcs reg (week4= strat1 strat2 strat3 week0 week1 week2 week3);
 fcs reg (week5= strat1 strat2 strat3 week0 week1 week2 week3 week4);
 fcs reg (week6= strat1 strat2 strat3 week0 week1 week2 week3 week4 week5);
```

```
 fcs reg (week7= strat1 strat2 strat3 week0 week1 week2 week3 week4 week5 week6);
 fcs reg (week8= strat1 strat2 strat3 week0 week1 week2 week3 week4 week5 week6 week7);
 fcs reg (week9= strat1 strat2 strat3 week0 week1 week2 week3 week4 week5 week6 week7
week8);
 fcs reg (week10= strat1 strat2 strat3 week0 week1 week2 week3 week4 week5 week6 week7
week8 week9);
 fcs reg (week11= strat1 strat2 strat3 week0 week1 week2 week3 week4 week5 week6 week7
week8 week9 week10);
 fcs reg (week12= strat1 strat2 strat3 week0 week1 week2 week3 week4 week5 week6 week7
week8 week9 week10 week11);
 mnar model(week0 week1 week2 week3 week4 week5 week6 week7 week8 week9 week10
week11 week12/ modelobs= (trt='1'));
run;


proc mixed data=final_imp(where=(avisitn=12));
by _imputation_;
 class trt(ref='1') strat1 strat2 strat3;
 model chg = trt strat1 strat2 strat3 base/ddfm=kenwardroger solution;
 lsmeans trt /pdiff cl;
 ods output lsmeans=lsmeans diffs=diffs;
run;


proc sort data=lsmeans;
by trt _imputation_ ;
run;
proc mianalyze parms=lsmeans;
modeleffects trt;
ods output ParameterEstimates=lsm;
by trt;
run;


proc sort data=diffs;
```

```
by _imputation_;
run;
proc mianalyze parms=diffs;
modeleffects trt;
ods output ParameterEstimates=dif;
run;
```

2. For secondary endpoint

```
proc mi data=vhcd_t out=imputed seed=128 nimpute=1000
 min = . . .. . 0 0 ;
class trt strat1 strat2 strat3 sex;
 var  trt strat1 strat2 strat3 sex week0 week24 ;
 fcs reg (week0= trt strat1 strat2 strat3  sex /details);/* if there is no missing for baseline, it will be ignored*/
 fcs reg (week24= trt strat1 strat2 strat3 sex week0 /details);
run;


proc mixed data=final_imp(where=(avisitn=24));
by _imputation_;
 class trt(ref='1') strat1 strat2 strat3;
 model chg = trt strat1 strat2 strat3 base/ddfm=kenwardroger solution;
 lsmeans trt /pdiff cl;
 ods output lsmeans=lsmeans diffs=diffs;
run;


proc sort data=lsmeans;by trt _imputation_ ;run;
proc mianalyze parms=lsmeans;
modeleffects trt;
ods output ParameterEstimates=lsm;
by trt;
run;
```

```
proc sort data=diffs;
by _imputation_;
run;
proc mianalyze parms=diffs;
modeleffects trt;
ods output ParameterEstimates=dif;
run;
```

Signature Page for TAK-062-2001_SAP_v4.0_03 Dec 2024_Final
Title: TAK-062-2001 SAP V4.0 03DEC2024

Approval	[redacted] Statistics 04-Dec-2024 02:24:58 GMT+0000

Approval	[redacted] Statistics 04-Dec-2024 03:32:59 GMT+0000

Document Number: TDN-000477385 v1.0