Title: Assessment of Immune Activation and Tolerance in Celiac Disease During Gluten Challenge

NCT Number: NCT03409796

SAP Approve Date: 3 May 2018

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

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

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

STATISTICAL ANALYSIS PLAN

STUDY NUMBER: TIMP-GLIA-5001

Assessment of Immune Activation and Tolerance in Celiac Disease During Gluten Challenge

PHASE 0

Version: Final

Date: 3 May 2018

Prepared by:

PPD

Based on:

Protocol Version: Amendment 1

Protocol Date: 19 December 2017

1.1 Approval Signatures

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

Study Title: Assessment of Immune Activation and Tolerance in Celiac Disease During Gluten Challenge

Approvals:

PPD

7 May 2018
Date

7 May 2018
Date

7 MAY 2018
Date

7 May 2018
Date

2.0 TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

3.0 LIST OF ABBREVIATIONS

APC	antigen-presenting cells
AE	adverse event
AUC	area under the curve
BMI	body mass index
CDSD	celiac disease symptom diary
CeD	celiac disease
CFR	Code of Federal Regulations
CRO	contract research organization
CV	coefficient of variation
DGP	anti-deamidated gliadin peptide
ECG	electrocardiogram
eCRF	electronic case report form
FDA	Food and Drug Administration
FFPE	formalin-fixed paraffin block
FSH	follicle-stimulating hormone
GCP	Good Clinical Practice
GFD	gluten-free diet
GLMM	generalized linear mixed model
IA	interim analysis
ICF	informed consent form
ICH	International Conference on Harmonisation
IEC	institutional ethics committee
IEL	intraepithelial lymphocyte
CCI	[redacted]
INF-γ	interferon-γ
IRB	institutional review board
LoQ	limit of quantification
mRNA	messenger RNA
MedDRA	Medical Dictionary for Regulatory Activities
MHC	major histocompatibility complex
MVCS	minutes of video with celiac symptoms
CCI	[redacted]
CCI	[redacted]
PT	preferred term
SAE	serious adverse event
SD	standard deviation
SOC	system organ class
SUSAR	suspected unexpected serious adverse reactions
TCR	T cell receptor
TEAE	treatment-emergent adverse event

tTG	anti-tissue transglutaminase
CCI	[redacted]
Vh:Cd	villus height to crypt depth ratio

4.0 OBJECTIVES

4.1 Primary Objectives

The primary objective of the trial is to characterize changes in gluten-specific T cells and pathology in the small intestine with specific focus on biomarkers likely to change with therapeutic celiac disease (CeD) treatment.

4.2 Secondary Objectives

The secondary objectives of this trial are as follows:

- To assess correlation between gluten-specific blood T cells and standard CeD histological assessments.
- To assess changes from Baseline in gluten-specific T cells in blood.

4.3 Exploratory Objectives

CCI

[redacted]

4.4 Study Design

This is a randomized, double-blind, 2-part gluten challenge trial in subjects with CeD who are HLA-DQ2.5 and/or HLA-DQ8 positive and have been on a gluten-free diet (GFD) for at least 6 months. Subjects will be enrolled until a maximum of approximately 20 subjects complete the gluten challenge and follow-up endoscopy.

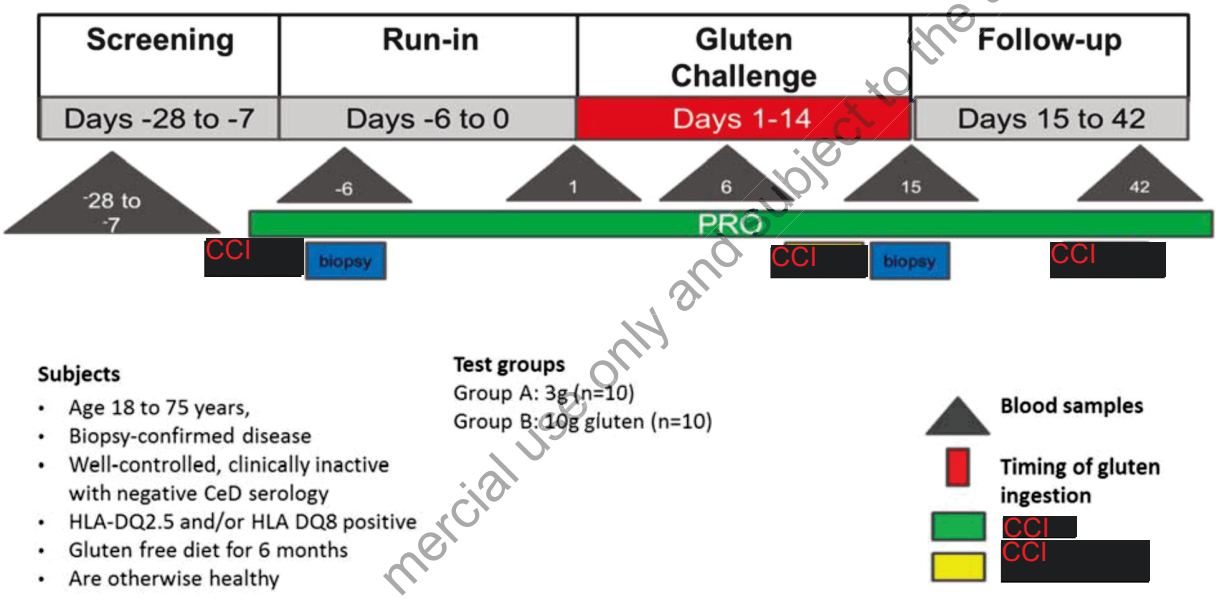
The trial will consist of 4 periods: Screening (Days -28 to -7), Run-in (Days -6 to 0), Gluten Challenge (Days 1-14, inclusive), and Follow-up (Days 15-42).

After signing the informed consent form (ICF), subjects will be enrolled and randomly assigned to 1 of 2 treatment groups: 3 g gluten/day or 10 g gluten/day. Subjects will be group randomized by site and will be enrolled into the two treatment groups in a 1:1 ratio. Both groups will be treated concurrently. Subjects receiving 10 g gluten/day will be able to reduce their dose to 3 g gluten/day after Day 3 if needed because of severity of symptoms. Dose reduction will be managed by an unblinded qualified staff member not directly involved in the treatment or clinical evaluation of the subjects. A complete gluten challenge requires that at least 12 of the 14 doses of gluten are taken before the follow-up endoscopy.

Subjects will be genotyped at Screening. If a subject has already been genotyped, results from previous testing may be used in lieu of genotyping at Screening. Symptoms will be measured

daily by subjects using the celiac disease symptom diary (CDSD). During Run-in, each subject will undergo a single, traditional endoscopy, and blood sampling to establish Baseline values. CCI will also be performed. During Gluten Challenge, all subjects will undergo periodic blood sampling. The day after completion of the gluten challenge, a second traditional endoscopy and biopsy will occur and subjects will undergo CCI Endoscopy will include video when feasible. During Follow-up, subjects will complete the CDSD, provide an additional blood sample, and undergo periodic CCI .

Figure 4.1 Study Schematic



CeD=celiac disease, CCI , CCI

5.0 ANALYSIS ENDPOINTS

5.1 Primary Endpoint

The primary endpoint of the trial is change from Baseline in small intestine histology based on standard CeD histological assessments of intraepithelial lymphocyte (IEL) counts and Vh:Cd measures.

5.2 Secondary Endpoints

Secondary endpoints are as follows:

- Correlation between gluten-specific blood T cells and standard CeD histological assessments.
- Changes from Baseline in gluten-specific blood T cells, based on functional assays and/or gluten-specific TCR staining.

5.3 Exploratory Endpoints

CCI

CCI

[redacted]

6.0 DETERMINATION OF SAMPLE SIZE

The sample size determination for this trial reflects several constraints. Despite having the primary analysis pooled across the 2 treatment arms, having the arms balanced is preferable so there must be an even number of subjects overall. The number of subjects was selected to detect a change in Vh:Cd in the interim analysis with 90% probability and at the final analysis with >99% probability. Additionally, it was selected to allow accurate estimation of the coefficient of variation for the secondary and exploratory endpoints, as well as to allow for comparing the 2 dosage groups.

Previous work estimates $\Delta Vh:Cd_{14}$ = -1.03 (0.96) (Vh:Cd change from Baseline after a 14-day gluten challenge) [1] and $\Delta \log_{10}(Tetramer)_6$ = 1.87 (1.39) (the log-10 fold change in tetramer binding 6 days into gluten challenge) [2]. Using these estimates, a single analysis with 12 subjects has 97% power to detect a change in Vh:Cd before and after gluten challenge, and >99% power to detect a change in T-cell marker tetramer binding. With the current group sequential design, then the interim analysis has 92% power of detecting a change in histology and an independent 98% power of detecting a change in T cell markers, for an overall power of 90% to detect both. There is >99% power to detect changes in both by the final analysis.

Uncertainty in the estimates of these 2 markers may lead to questioning the accuracy of these measurements. If it is assumed that the SD of $\Delta Vh:Cd_{14}$ and $\Delta \log(Tetramer)_6$ is 50% greater than is reported in the literature, the power of this trial would be decreased. In this scenario, the trial has a 61% power to detect a histological change in the interim analysis and 91% power to detect a histological change overall. In addition, the probability of detecting a change in T cell markers is 75% at the interim analysis and 99% by the final analysis.

7.0 METHODS OF ANALYSIS AND PRESENTATION

7.1 General Principles

Baseline values are defined as the last observed value before the first dose of gluten.

All statistical analyses will be conducted using SAS® Version 9.1, or higher.

All statistical tests, and resulting P-values will be reported as 1-sided and will be assessed at α=0.05 significance level unless otherwise stated. Confidence intervals will be reported as both 1-sided and 2-sided at α=0.05 significance level unless otherwise stated. P-values will be rounded to 3 decimal places for outputs.

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. Confidence intervals about a parameter estimate will be presented using the same number of decimal places as the parameter estimate.

Where appropriate, variables will be summarized descriptively by study visit based on observed value and by change from Baseline.

For the categorical variables, the count and proportions of each possible value will be tabulated by treatment group and overall. The denominator for the proportion will be based on the number of subjects who provide non-missing responses to the categorical variable.

For continuous variables, the number of subjects with non-missing values, mean, standard deviation (SD), minimum, first quartile, median, third quartile, and maximum values will be tabulated. In some instances, the $\log_{10}$-transformed values will be reported as well. For differences from Baseline, the coefficient of variation (CV), defined as the standard deviation divided by the mean, will be reported as well. Boxplots of measurements at different timepoints will also be generated.

Screen failure subjects will be presented separately via listings.

7.1.1 Study Definitions

7.1.2 Definition of Study Days

Study Day 1 is defined as the date on which a subject is administered their first dose of the study product. Other study days are defined relative to the Study Day 1, with Day 0 being the day prior to Study Day 1, Day -1 being two days prior to Study Day 1, and so forth.

Table 7.1 Definition of Study Windows

Window	Start Day	End Day
Screening	-28	-7
Run-in	-6	0
Gluten Challenge	1	14
Follow-up	15	42

7.2 Analysis Sets

7.2.1.1 Safety Set

The safety set will consist of all subjects who are enrolled and receive at least 1 dose of gluten. Subjects in this analysis set will be used for demographic, Baseline characteristics, and safety summaries.

7.2.1.2 Biomarker Analysis Set

The biomarker analysis set will consist of all subjects who complete the gluten challenge and have all the Vh:Cd measures scheduled in the protocol taken and be non-missing.

7.3 Disposition of Subjects

Subject disposition at the end of the study will be summarized by treatment group and overall. A table will display the number of subjects that:

- Enrolled
- Are in each analysis set (Safety and Biomarker Analysis)
- Completed the study per protocol
- Had a Dose Reduction, along with the reason for dose reduction
- Discontinued the study, along with the reason for discontinuation

7.4 Demographic and Other Baseline Characteristics

Baseline demographics will be summarized for all patients in the safety set and biomarker analysis set. Demographic data will be summarized by treatment arm and overall. Baseline demographic data to be evaluated will include age at date of informed consent, sex, ethnicity, race, height, and weight. Patient enrollment by site will also be summarized by treatment arms and overall. No inferential statistics will be generated. Demographic data will also be presented in a by-patient listing.

7.5 Medical History and Concurrent Medical Conditions

General medical history and concurrent medical conditions will be listed for all patients.

7.6 Medication History and Concomitant Medications

General medication history and concomitant medications will be listed for all patients.

7.7 Gluten Exposure and Compliance

Subjects who do not comply with their gluten dosage will be listed with the daily gluten amount taken. Subjects who do not complete the gluten challenge will be excluded from further analysis. Subjects who started at the 10g gluten dose and had their dosage reduced to 3g will be included in the overall analysis, but excluded from analyses by treatment arm.

7.8 Efficacy Analysis

{Not applicable}

7.9 Pharmacokinetic/Pharmacodynamic Analysis

{Not applicable}

7.10 Pharmacokinetic Analysis

{Not applicable}

7.10.1 Pharmacodynamic Analysis

{Not applicable}

7.11 Other Outcomes

7.11.1 Primary Biomarker Endpoint

The primary endpoint of the trial is the change from Baseline in small intestinal histology. This is based on the following measures that are typically used for diagnosing CeD:

- The ratio of villous height to crypt depth (Vh:Cd).
- Intraepithelial lymphocytes (IEL) counts

Smaller values of Vh:Cd and greater IEL counts indicate more extreme celiac disease symptoms. The purpose of the primary analysis is to confirm a response to gluten by measuring the change in these measures. Only the change in Vh:Cd will be used to confirm a gluten response, and the change in IEL counts is assessed as an alternative measure for future work. Any statistically significant change in Vh:Cd will be considered confirmation of gluten response.

7.11.1.1 Vh:Cd

Vh:Cd will be measured during Run-in ($Vh:Cd_R$) and Day 15 ($Vh:Cd_{15}$). The Vh:Cd change from Baseline ($\Delta Vh:Cd = Vh:Cd_{15} - Vh:Cd_R$) will be computed for all subjects in the Biomarker Analysis Set. These values are continuous variables and tabulated as laid out in General Principles [7.1].

ΔVh:Cd is assumed to follow a normal distribution, and a 1-sided paired t-test will be used to compare Baseline and follow-up Vh:Cd measures. The normality assumption will be checked using the Shapiro-Wilk test. If the data is found to not be normal at α=0.05, a 1-sided Wilcoxon signed-rank test will be used instead to compare $Vh:Cd_R$ and $Vh:Cd_{15}$. The null and alternate hypotheses for Vh:Cd test are below:

H_0: Baseline Vh:Cd ($Vh:Cd_R$) ≤ Follow-up Vh:Cd ($Vh:Cd_{15}$); $\Delta Vh:Cd \geq 0$
H_A: Baseline Vh:Cd ($Vh:Cd_R$) > Follow-up Vh:Cd ($Vh:Cd_{15}$); $\Delta Vh:Cd < 0$

7.11.1.2 IEL Counts

Statistics on IEL counts are not being used for decision making purposes and no alpha is being spent on their analysis. IEL counts will be measured on Run-in (IEL_R) and Day 15 (IEL_{15}). The IEL change from Baseline ($\Delta IEL = IEL_{15} - IEL_R$) will be computed for all subjects in the Biomarker Analysis Set. These values are continuous variables and tabulated as laid out in General Principles [7.1].

IEL counts are assumed to follow a Poisson distribution. The Poisson distribution assumption will be checked using the Kolmogorov-Smirnov test. If the data is found to not be Poisson at α=0.05, a 1-sided Wilcoxon signed-rank test will be used instead to compare IEL_R and IEL_{15}. Otherwise a Poisson GLMM will be fitted to the data, where IEL measurements are grouped by subject (the random effect) and the change in IEL counts is the fixed effect. The formula for this model is below:

$$\log[E(IEL|Day, Subject)] = \beta_0 + Day \times \beta + Subject \times u$$

Here, *IEL* is the measured IEL count, β_0 is the intercept, *Day* is an indicator for day 15, β is the measured fixed effect of how much IEL counts change, *Subject* is a grouping factor for measurements by subject, and u is the subject-specific random effect. A Wald test will be performed on the change in IEL counts.

The null and alternate hypotheses for IEL are below:

H_0: Baseline IEL counts (IEL_R) ≥ Follow-up IEL counts (IEL_{15}); $\Delta IEL \leq 0$
H_A: Baseline IEL counts (IEL_R) < Follow-up IEL counts (IEL_{15}); $\Delta IEL > 0$

7.11.2 Secondary Biomarker Endpoints

The secondary endpoints involve measuring gluten-specific blood T cells via the ELISpot assay or tetramer/dextramer antigen staining. These are the biomarkers being evaluated to augment or replace histological markers in future CeD studies. Each T cell marker assay will be tested by itself but using the same procedure as laid out below.

The T cells will be measured at Run-In (T_R), Day 1 (T_1), Day 6 (T_6) and Day 15 (T_{15}). For each T cell assay, values below the limit of quantification (LoQ) will be replaced with LoQ/2. The T cell log fold change from Baseline at Day 6 ($\Delta T = \log_{10}(T_6/T_B)$) will be computed for all subjects in the Biomarker Analysis Set. T_B is the last observed T cell value before gluten exposure.

The T cell assays measure staining intensity and are assumed to follow a log-normal distribution (and by extension, that ΔT follows a normal distribution). In addition to having the T cells tabulated as laid out in General Principles [7.1], the $\log_{10}$-transformed values will be tabulated as well. The distributional assumptions will be tested by log-transforming the data and using the Shapiro-Wilk test. If the data is found to be not log-normal at α=0.05, nonparametric tests will be used instead.

7.11.2.1 Correlation

T cell measurements taken before the first dose of gluten will be correlated with Baseline Vh:Cd and IEL counts, and T cell measurements taken after the first dose of gluten will be correlated with day 15 Vh:Cd and IEL counts using the Spearman correlation.

Additionally, ΔT will be correlated with ΔVh:Cd and ΔIEL. If the assumptions that ΔVh:Cd and ΔT follow a normal distribution are not rejected, then they are correlated with a Pearson correlation. Otherwise, a Spearman correlation will be used. ΔT will be correlated with ΔIEL with a Spearman correlation as well.

7.11.2.2 Change from Baseline

Depending upon availability, one of these T cell markers will be the main secondary endpoint, with preference going to ELISpot, then tetramer, then dextramer antigen staining. If the assumption that T cell markers follow a log-normal distribution is not rejected, a pairwise 1-sided t test will be used to compare Baseline and Day 6 $\log_{10}$-transformed T cell measurements. If the assumption is rejected, then a Wilcoxon signed-rank will be used instead. The null and alternate hypotheses of this test are below:

H_0: Baseline T cells ($T_{Baseline}$) ≥ Day 6 T cells (T_6); $\Delta T \leq 0$
H_A: Baseline T cells ($T_{Baseline}$) < Day 6 T cells (T_6); $\Delta T > 0$

7.11.2.3 Linear mixed model

A linear mixed model will also be used to compare pre-dose to post-dose $\log_{10}$ T cell marker measures; this test will use all available T cell marker measures for a given assay. This mixed effect model will follow the following formula:

$$T = \beta_0 + Dose \times \beta + Subject \times u + \varepsilon$$

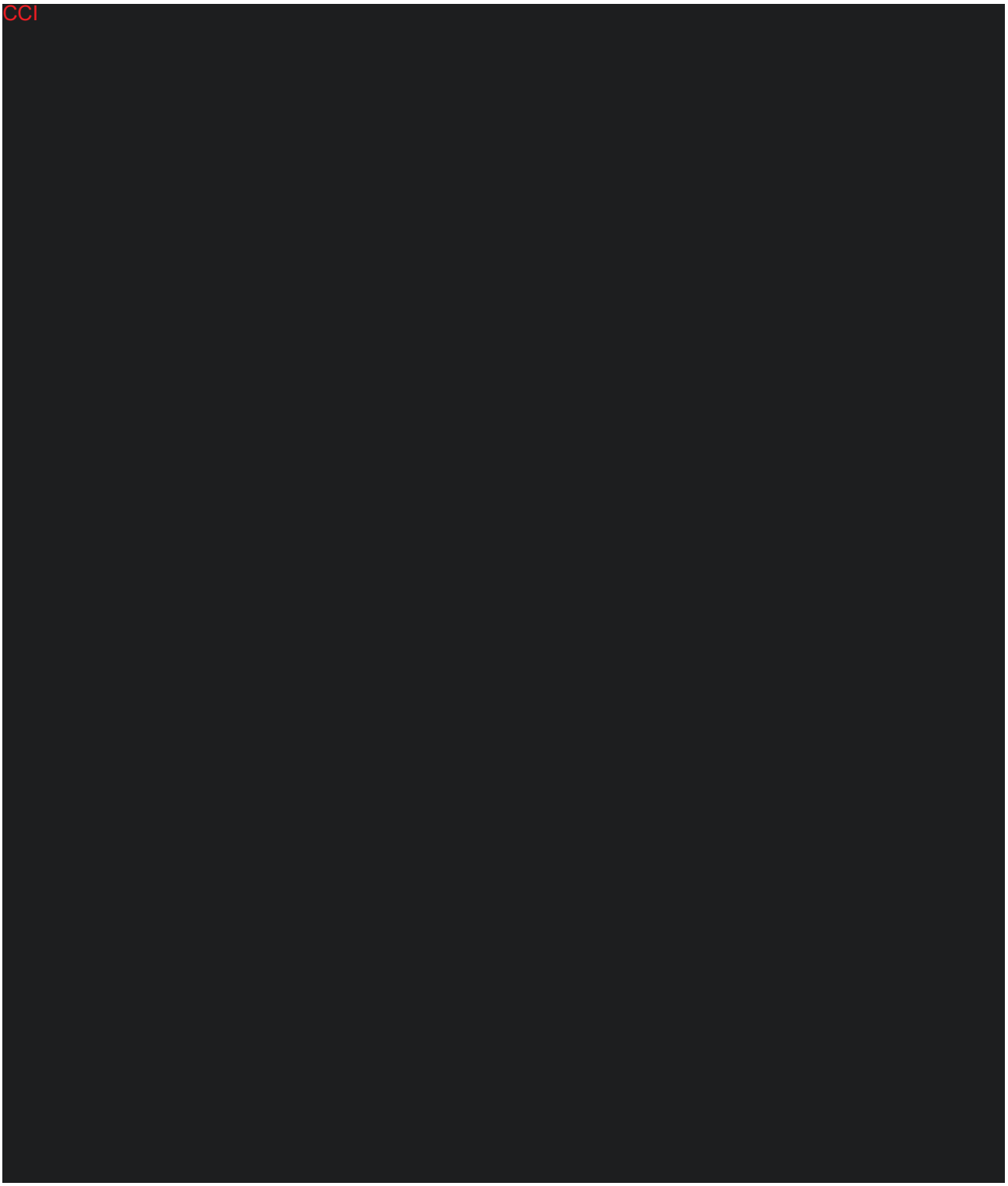
Here, T is $\log_{10}$ T cell measurement, β_0 is the intercept, *Dose* is the indicator of post-gluten-exposure, *Subject* is a grouping factor for measurements by subject, u is the subject-specific random effect, and ε is noise. The value of β will be reported along with standard error and p-value.

7.11.3 Exploratory Biomarker Endpoints

Due to the exploratory nature of the following endpoints, their distributions are not known. Each biomarker and its change from Baseline should be tabulated and plotted as in General Principles [7.1]. Some biomarker endpoints may also need to be transformed prior to analysis.

CCI

CCI



CCI

7.12 Safety Analysis

Safety evaluations will be based on the incidence, severity, type of AEs, clinically significant changes or abnormalities in the patient's physical examination, vital signs, and clinical laboratory results.

These analyses will be performed using the Safety Set.

7.12.1 Adverse Events

AEs will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA). Treatment emergent is defined as any AE that occurs after administration of the first dose of gluten and up through the end of the follow-up period, any event that is considered gluten-related up through 30 days post-last-dose, or any event that is present at Baseline but worsens in severity after Baseline up through 30 days post-last-dose.

7.12.2 Clinical Laboratory Evaluations

Not applicable; routine laboratory evaluations will be conducted only for the purpose of screening and enrolling subjects.

7.12.3 Vital Signs

Baseline, post-dose, and changes from Baseline in vital sign measurements will be summarized. All vital sign data will be provided in data listings.

7.12.4 12-Lead ECGs

{Not applicable}

7.12.5 Other Observations Related to Safety

{Not applicable}

7.13 Interim Analysis

There is one interim analysis (IA) planned for this study which occurs after 12 out of a possible 20 subjects have completed the study. The purpose of the interim analysis is to determine if an additional 8 patients will be enrolled in the study. The IA consists of the three stages (see flowchart):

1. The first stage confirms a gluten response with a standard histological measure (Vh:Cd) [7.11.1.1]. If a gluten response is detected, proceed to the second stage. If a gluten response is not detected, proceed to stage 3a. The table below contains the critical values for Vh:Cd.

Table 7.2 Vh:Cd critical values

Analysis	Subjects	Critical Value	α	Cumulative α
Interim	12	0.0240	0.0240	0.0240
Final	20	0.0373	0.0260	0.0500

2. If a histological change is observed, the second stage tests whether the gluten response is observed via on of the T cell markers [7.11.2.2], with preference going to ELISpot (if available), then tetramers and then dextramers. If the gluten response is detected with the T cells, stop the study. If the gluten response is not detected with the T cells, proceed to stage 3b. The table below contains the critical values for the T cells.

Table 7.3 T cell critical values

Analysis	Subjects	Critical Value	α	Cumulative α
Interim	12	0.0179	0.0179	0.0179
Final	20	0.0417	0.0321	0.0500

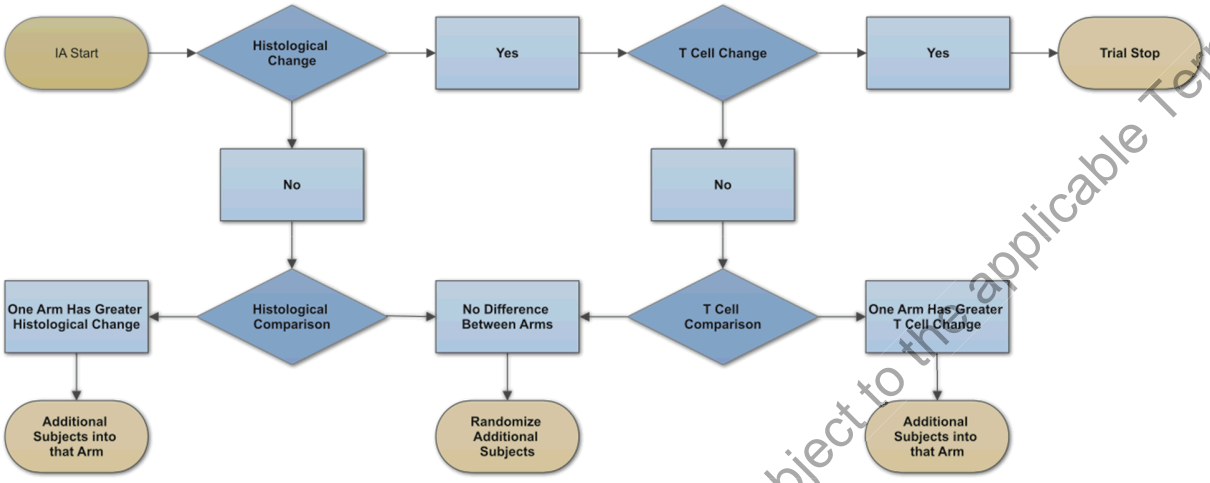
3. If a histological change is not observed or is not detected with the T cell marker, then more subjects will be needed. The third stage compares the two treatment arms.
    a. If a histological change was not observed, then the third stage compares the two arms using a standard histological measure (Vh:Cd).
    b. If a T cell marker change was not observed, then the third stage compares the two arms using a T cell marker.

The purpose of the third stage is to test whether a larger gluten response is observed in either of the two treatment arms. This will be performed using a two-sided two-sample t test comparing ΔVh:Cd$^{(3)}$ and ΔVh:Cd$^{(10)}$ (the change in Vh:Cd in the 3g and 10g dosage arms, respectively) or ΔlogT$^{(3)}$ and ΔlogT$^{(10)}$ (the log fold change of the T cells in the 3g and 10g dosage arms, respectively). If ΔVh:Cd or ΔlogT were found to not follow normal distributions previously, then a Mann-Whitney U test will be used instead. The null and alternate hypotheses of this test are below:

H_0: The change from Baseline of the marker is the same in both groups ($\Delta X^{(3)} = \Delta X^{(10)}$)
H_A: the change from Baseline of the marker is unequal ($\Delta X^{(3)} \neq \Delta X^{(10)}$)

If this test is significant at α=0.05, then subjects will be enrolled into the treatment arm with greater gluten response (more negative ΔVh:Cd or more positive ΔlogT**)**. If this test is not significant, then patients will be enrolled equally into the two dosage groups after the interim analysis.

Figure 7.1 Interim Analysis Flowchart



7.14 Changes in the Statistical Analysis Plan

CCI analysis endpoints have been clarified to used CDSD.

8.0 REFERENCES

[1] D. Leffler, D. Schuppan, K. Pallav, R. Najarian, J. D. Goldsmith, J. Hansen, T. Kabbani, M. Dennis and C. P. Kelly, "Kinetics of the histological, serological and symptomatic responses to gluten challenge in adults with coeliac disease," *Gut,* vol. 62, no. 7, pp. 996-1004, 2013.

[2] V. K. Sarna, G. I. Skodje, H. M. Reims, L. F. Risnes, S. Dahal-Koirala, L. M. Sollid and K. E. Lundin, "HLA-DQ:gluten tetramer test in blood gives better detection of coeliac patients than biopsy after 14-day gluten challenge," *Gut,* vol. 0, pp. 1-8, 18 August 2017.