Statistical Analysis Plan: I6T-MC-AMAC (Version 3)

A Phase 2, Multicenter, Randomized, Double-Blind, Parallel, Placebo-Controlled Study of LY3074828 in Subjects with Moderate to Severe Ulcerative Colitis

NCT02589665

Approval Date: 29-Jan-2018

1. Statistical Analysis Plan:
I6T-MC-AMAC: A Phase 2, Multicenter, Randomized, Double-Blind, Parallel, Placebo-Controlled Study of LY3074828 in Subjects with Moderate to Severe Ulcerative Colitis

LY3074828 Ulcerative Colitis

Study I6T-MC-AMAC is a Phase 2, multicenter, randomized, double-blind, placebo-controlled trial of LY3074828 in subjects with moderate to severe ulcerative colitis.

Eli Lilly and Company
Indianapolis, Indiana USA 46285
Protocol I6T-MC-AMAC
Phase 2

Statistical Analysis Plan electronically signed and approved by Lilly on date provided below.

APPROVAL

A Phase 2, Multicenter, Randomized, Double-Blind, Parallel, Placebo-Controlled Study of LY3074828 in Subjects with Moderate to Severe Ulcerative Colitis

PROTOCOL I6T-MC-AMAC

STATISTICAL ANALYSIS PLAN

Version 3.0

PPD

26 Jan 2018

Date

PPD

26 Jan 2018

Date

PPD
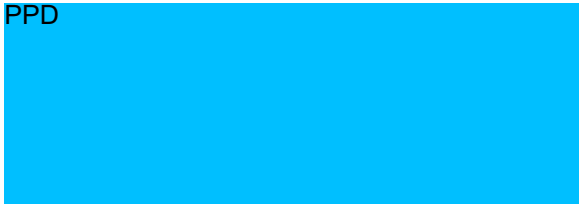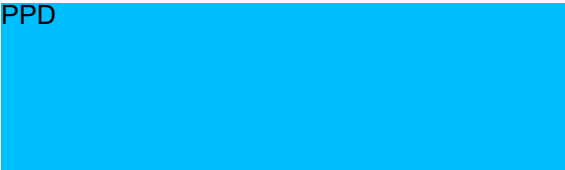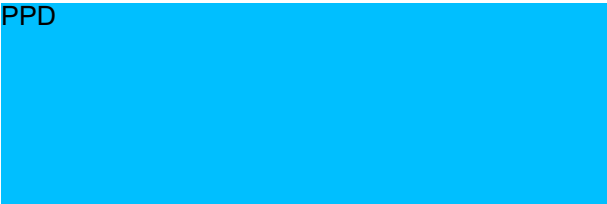
26/Jan/2018.

Date

PPD
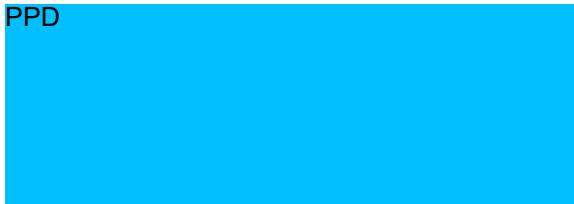
29 / Jan / 2018

Date

2. Table of Contents

2.1. List of Appendices

Appendix A. The UCEIS

Appendix B. IBDQ Missing Data Handling Algorithm

Appendix C. Histopathology

2.2. List of Tables

2.3. List of Figures

Figure 4   Flow Chart of ADA Sample Assessment with Clinical Interpretation

2.4. List of Abbreviations

ADA	Anti-Drug Antibodies
AE	Adverse Event
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Classification
BMI	Body Mass Index
CSR	Clinical Study Report
eCRF	Electronic Case Report Form
ECG	Electrocardiogram
HLT	High Level Term
IBDQ	Inflammatory Bowel Disease Questionnaire
IRR	Infusion Related Reactions
ITT	Intent to Treat
IC	Informed Consent
IV	Intravenous
IWRS	Interactive Web-Response System
LLN	Lower Limit of Normal
LLT	Lowest Level Term
LSS	Lilly Safety System
LOCF	Last Observation Carried Forward
mBOCF	modified baseline observation carried forward
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed-Effect Model Repeat Measurement
MRD	Minimum Required Dilution

Msec	Millisecond
Nab	Neutralizing Antibody
NRI	Non-Responder Imputation
OLE	Open-Label Extension
PGI-I	Patient's Global Impressions of Improvement
PGI-S	Patient's Global Impressions of Severity
PK/PD	Pharmacokinetics/Pharmacodynamics
PT	Preferred Term
Q4W	Every 4 Weeks
Q12W	Every 12 Weeks
QNS	Quantity Not Sufficient
QTcF	Corrected QT (Fridericia)
QTcLCTPB	Corrected QT (large clinical trial population based)
RHI	Robarts Histopathology Index
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SC	Subcutaneous
SF36	36-Item Short Form Health Survey
SI	International System of Units
SMQ	Standardised Medical Queries
SOC	System Organ Class
TDM	Therapeutic Drug Monitoring
TEAE	Treatment-Emergent Adverse event
UC	Ulcerative Colitis
UCEIS	Ulcerative Colitis Endoscopic Index
WHO	World Health Organization

3. Revision History

SAP Version 2 was approved prior to first unblinding.

SAP Version	Version Date	Comments
Version 1.0	20JUNE2016	Version 1.0 was approved prior to unblinding.
Version 2.0	02JUNE2017	Version 2.0 was approved prior to unblinding.
Version 3.0	11JAN2018	Version 3.0 was approved post unblinding of study.

Revisions Post Final SAP Version 1.0 20JUNE2016

Section	Summary of Changes
2.4	Added Informed Consent (IC)
3	Added approval date for SAP v2.
4.3	Added UCEIS Reference
5.1	• Modified definition for prior biologic therapy. • Added Maintenance Rescue criteria at w 52 and beyond • End of maintenance dosing period changed from Week 52 to Week 104. • Study schematics updated to reflect the longer study duration. • Modified the Extension Induction dose to 1000 mg IV from 600mg IV.
5.3	The dose for Extension Period – Induction portion of the study changed from 600 mg to 1000 mg
6.1	• Updated parties responsible for TFL generation. • Added Follow up Period to list of study periods. • Clarified wording around study periods (from treatment periods). • Doses for Extension Period – Induction in Table 1 changed from 600 mg to 1000 mg

Section	Summary of Changes
	- Clarified Dose labels in Table 1 to not include combined period labels. - Description of how to reset placebo group baseline when entering the extension period modified. "assessment prior to the date of first injection" changed to "assessment on or prior to the date of first injection" - Added wording to clarify that not all analyses mentioned in the SAP will necessarily be created in a static display.
6.2	Added flexible wording around the possibility of a Per Protocol population.
6.4	Wording added to reference the new Appendix B (IBDQ Missing Data Algorithm)
6.4.2	Clarified use of mBOCF for health outcomes analyses is limited to PGI-S and IBDQ.
6.4.3	Added section for LOCF.
6.4.4	Removed "follow-up emergent" AE language and section number change.
6.5	Changed the interaction p-value criteria to 0.10 from 0.05.
6.6	Changed references for the graphical procedure.
6.8	Changed randomization date to informed consent date for calculation of age
6.10	- Added wording around screening period concomitant medication summary. - Changed Concomitant Medication start date definition to Informed Consent date from first dose date. - Removed wording describing an analysis of prior Concomitant Medications and current Concomitant Medications.
6.11.5	- Added additional exploratory efficacy endpoints which were not previously described in the protocol. - Added Reference to UCEIS. - Added a definition of time to response and duration of response. - Added censoring rules for time to event endpoints.
6.12	Added "6.12.1 Secondary Health Outcomes Endpoints" heading.
6.14	Added wording regarding the formal statistical testing of safety data.

Section	Summary of Changes
	• Clarifying wording added and certain analyses removed per Safety Analytics guidance. • Description of how to reset placebo group baseline when entering an extension period changed from "prior to the first injection" to "on or prior to the first injection"
6.14.1	• Clarified the periods for which labs will be evaluated. • For each period, clarified the method for calculating duration of exposure to include the disposition event dates. • Removed reference to analysis of combined periods.
6.14.2	• Added wording around screening period AE summary. • Pre-specified creation of a by patient AE listing removed. • Pre-specified serious adverse events wording removed. • Clarified use of Prior Biologic eCRF question in subgroup identification. • Added AEs leading to study drug discontinuation to the criteria for notable patients for whom narratives will be generated. This is per the Lilly recommended process.
6.14.2.2	• Changed heading from "Analysis of Potential Hypersensitivity, Anaphylaxis, and Infusion Related Reactions" to "Hypersensitivity Reactions". • Changed "Analysis 1" to "Time Period A". • Changed "Analysis 2" to "Time Period B". • Updated definitions and summaries of hypersensitivity reactions to align with current standards. • Other wording clarifications to align with current standards.
6.14.2.3	• Changed heading from "Analysis of Administration Site Reactions" to "Analysis of Injection Site Reactions". • Updated summaries of injection site reactions to align with current standards.
6.14.3	Added wording around the analysis of AEs leading to discontinuation (other significant AEs).

Section	Summary of Changes
6.14.4	• Clarified that reference limits would be based on limits provided by the performing laboratory. • Removed wording about P-values and confidence limits not being included in box plot summaries.
6.14.4.1	• Changed "Covance" to "performing laboratory". • Updated the definition and summaries of hepatic laboratories to align with current standards
6.14.5	• Updated wording around Periods for Vital sign summaries. • Removed wording about P-values and confidence limits not being included in box plot summaries.
6.14.6	Clarified wording around ECG details. No details are collected.
6.14.7	Added Serum NGAL and MMP9.
6.14.8.4	Updated summaries of immunogenicity data to align with current standards
6.14.9	Updated details of the Geboes Scoring and RHI scoring methods for Histopathology analysis.
6.15.1	• Clarified the ATC drug code details for concomitant and prior biologic based analyses. • Added subgroup definition and details for patients receiving a TDM adjustment.
6.15.2	Update subgroup analyses of safety to indicate that they may be done
6.16	Added Table 9 to define important protocol violations
6.18	Added wording around the changes to the protocol specified analysis of concomitant medications.
8.	• Added additional Multiple Testing and UCEIS Reference
9.	• Added Appendix A-UCEIS • Added Appendix C- Histopathology
Throughout	• Changed 'baseline period' to 'screening period'. • Template language describing the program level statistical analysis removed.

Revisions Post Final SAP Version 2.0 02JUNE2017

Section	Summary of Changes
1	Updated titles for blinded statisticians
2.4	Added Open-Label Extension (OLE)
3	Added approval date for SAP v3 and summary of changes table post final SAP Version 2.
5.1	Added country specific addendum details.
6.1	Update TFL presentation for percentage counts. Clarified wording around displays presented for CSR and static displays.
6.2	Insertion of wording for visualization tools (e.g. Spotfire).
6.4	Added reference to missing partial end dates for important programmable deviations.
6.10	Removed paragraph summarizing data based on medical monitor identifying corticosteroids.
6.11.5	Added additional exploratory efficacy endpoint for Histologic Remission.
6.12.1.1	Added analysis details for the SF36 (36-Item Short Form Health Survey score).
6.13	Updated timing and unblinding details for PK/PD analyses.
6.14	Updated pairwise comparisons of LY3074828 for maintenance phase.
6.14.8.3	Updated definition of ADA Inconclusive status.
6.14.9	Added proposed definition of Histologic Remission and further definitions for exploratory purposes and analyses relating to Histologic Remission. Clarified additional wording.
6.14.5	Clarified baseline definition for vital signs.
6.16	Added imputation of prior medications for programmable protocol violations. Updated Table 9 to define important programmatic protocol violation checks.
6.17	Updated to include details on the timing of the Primary Analysis. Created subsections to describe the primary and interim analyses.
9	Updated title for table 12.

4. Study Objectives

4.1. Primary Objective

The primary objective of this study is to test the hypothesis that treatment with LY3074828 is superior to placebo in inducing clinical remission at Week 12 in patients with moderate to severe ulcerative colitis (UC).

4.2. Secondary Objectives

The secondary objectives are:

1. To evaluate the safety and tolerability of treatment with LY3074828.

2. To evaluate the efficacy of treatment with LY3074828 in inducing a clinical response at Week 12.

3. To evaluate endoscopic remission at Week 12 and Week 52.

4. To evaluate the effect of maintenance treatment with LY3074828 on the durability of clinical remission, endoscopic remission, and clinical response at Week 52.

5. To evaluate the effect of LY3074828 on health outcomes/quality of life measures (Inflammatory Bowel Disease Questionnaire score (IBDQ), 36-Item Short Form Health Survey score (SF-36) and Patient's Global Impressions of Severity (PGI-S) score, and Patient's Global Impressions of Improvement (PGI-I) score).

6. To characterize the pharmacokinetic (PK) profile of LY3074828.

4.3. Exploratory Objectives

The exploratory objectives are:

7. To assess the change from baseline in Ulcerative Colitis Endoscopic Index of Severity (UCEIS) [4] (completed by central reader only) scores at Weeks 12 and 52.

8. To assess the change from baseline in the partial Mayo score at various times during induction and maintenance.

9. To determine the relationship of LY3074828 exposure levels to clinical endpoints and biomarkers.

10. To determine the change from baseline in the biomarkers C-reactive protein (CRP) and fecal calprotectin.

11. To explore the development of any anti-LY3074828 antibodies that are formed (NGAL, MMP-9n) and their effect on safety, PK, and pharmacodynamics (PD) of LY3074828.

12. To explore the effect of additional (unblinded) dosing on clinical response in patients who do not meet responder criteria at Week 12.

13. To assess the severity of the disease by histopathological images using the Robarts Histopathology Index (RHI)

5. Study Design

5.1. Summary of Study Design

Study I6T-MC-AMAC is a multicenter, randomized, double-blind, parallel-arm, placebo-controlled trial in which approximately 240 patients will be randomized.

Patients must have moderate or severe UC (defined as a Mayo score of 6 to 12 with an endoscopic subscore ≥2). Approximately two-thirds (~160) of the patients randomized to study treatment must have been exposed to at least 1 previous biologic therapy (received treatment with 1 or more agents such as tumor necrosis factor [TNF] antagonists, vedolizumab, or experimental UC biologic therapy), and approximately one-third (~80) of the patients will be naive to biologic therapy.

Screening Period: Patients will be evaluated for study eligibility ≤28 days before the baseline visit.

At the baseline visit, patients who fulfill the eligibility criteria will be randomized equally to 1 of 4 induction treatment arms (LY3074828 50 mg, LY3074828 200 mg, LY3074828 600 mg, and Placebo).

Induction Dosing Period (Induction Period): A 12-week induction period is designed to establish the efficacy and safety of LY3074828 administered intravenous (IV) at Weeks 0, 4, and 8. Patients will be stratified across the treatment arms on the basis of previous exposure to biologic therapy for treatment of UC. Patients who discontinue the study for any reason before the end of the induction period will complete the early study discontinuation follow-up visit.

Maintenance Dosing Period (Maintenance Period): The maintenance period is designed to explore the safety and durability of clinical responses and remissions to treatment with 200-mg LY3074828 administered subcutaneously (SC) every 4 weeks (Q4W) or every 12 weeks (Q12W). Patients defined as having clinical responses at Week 12 will continue study participation in the maintenance period up to Week 104. Patients who do not meet clinical response criteria at Week 12 will have the option to continue in a study Extension Period or discontinue from the study.

Responding patients who have received LY3074828 in the induction period will be re-randomized to 1 of 2 LY3074828 maintenance treatment arms (200 mg SC Q4W or 200 SC mg Q12W); these patients will be stratified according to their Week-12 remission status. Any responding patients in the placebo arm will remain on placebo.

After Week 52, subjects who experience worsening of UC (partial Mayo score of 7 or more) may receive rescue treatment with LY3074828 200 mg SC Q4W.

Extension Period- Induction: Patients who complete the study induction period (through Visit 7) but do not have a clinical response may choose to participate in the study extension period following consultation with, and at the discretion of, the investigator.

During the extension period induction, all patients are planned to receive 1000 mg LY3074828 IV administered at Extension Weeks 0, 4, and 8 (the dose level and/or frequency of dosing may

be reduced based on review of Study AMAC trial data). Patients should complete all assessments according to the extension period schedule of events and should remain on permitted UC concomitant medication according to Section 9.9 of the protocol.

<u>Extension Period- Maintenance</u>: Patients who have a clinical response (Section 6.11.3.1) at Extension Visit 6 will have the opportunity to continue on extension period maintenance therapy, while non-responders at Extension Visit 6 will be discontinued from the study.

Patients who continue into the extension period maintenance treatment are planned to receive unblinded 200 mg LY3074828 administered SC Q4W, should complete all assessments according to the extension period schedule of events in Attachment 2 of the protocol, and should remain on permitted UC concomitant medication according to Section 9.9 of the protocol. The dose level and frequency of dosing may be reduced based on review of Study AMAC trial data.

An addendum has been planned for sites that have subjects that need to extend treatment in study AMAC in order to have the opportunity to participate in an OLE (Open Label Extension) study (this is Country specific). The duration of the maintenance period and extension maintenance period will be extended to provide for up to an additional 52 weeks of therapy so that subjects can continue to receive study drug until the OLE Study becomes available. Once an OLE study becomes available, subjects will be asked to consider participation in the OLE or be discontinued from AMAC. Patients will continue to be followed for safety and clinical benefit during this additional year of therapy and in addition the sponsor conducts quarterly trial level safety reviews (TLSRs) during the study. No TFLs or further analyses are planned for the addendum.

Figure 1 and Figure 2 show the study design.

Figure 1   Study Schematic (Induction and Maintenance Periods)
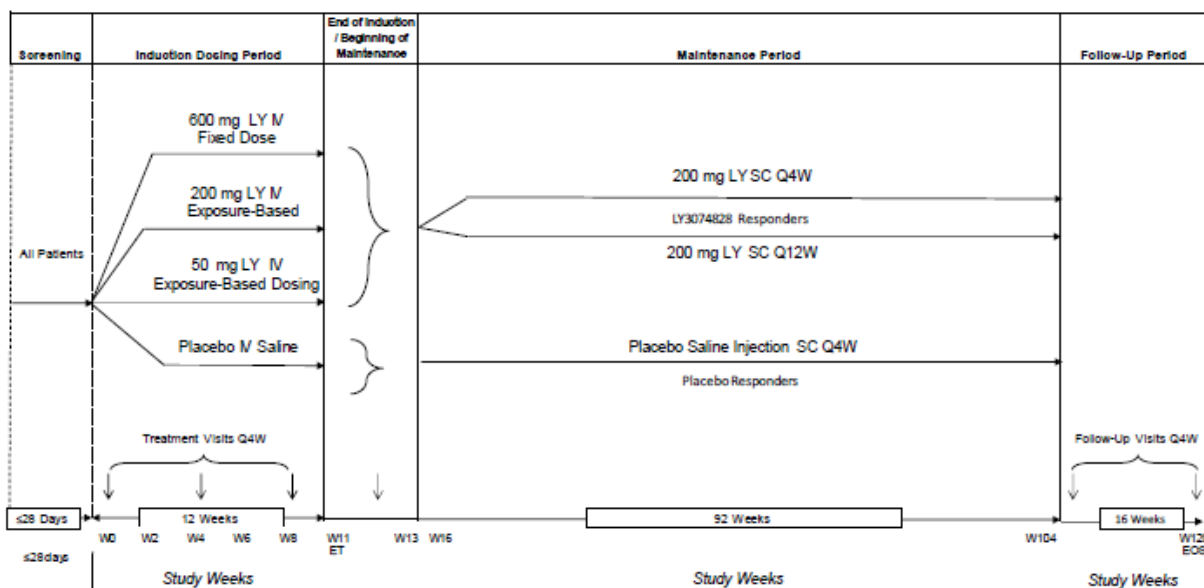
Abbreviations: IV = intravenous; LY = LY3074828; n = number of subjects; Pbo = placebo; Q4W = every 4 weeks; Q12W = once every 12 weeks; SC = subcutaneous; W = study week.

Note: Induction nonresponders can consider participation in the study extension period (Section 7.1.1 of the protocol).

Note: Exposure Based Dosing is synonymous with Therapeutic Dose Monitoring (TDM)

Figure 2   Study Schematic (Extension Period)

Abbreviations: EOS = end of study; ET = early termination; EW = extension week; IV = intravenous; LY = LY3074828; Q4W = every 4 weeks; SC = subcutaneous.
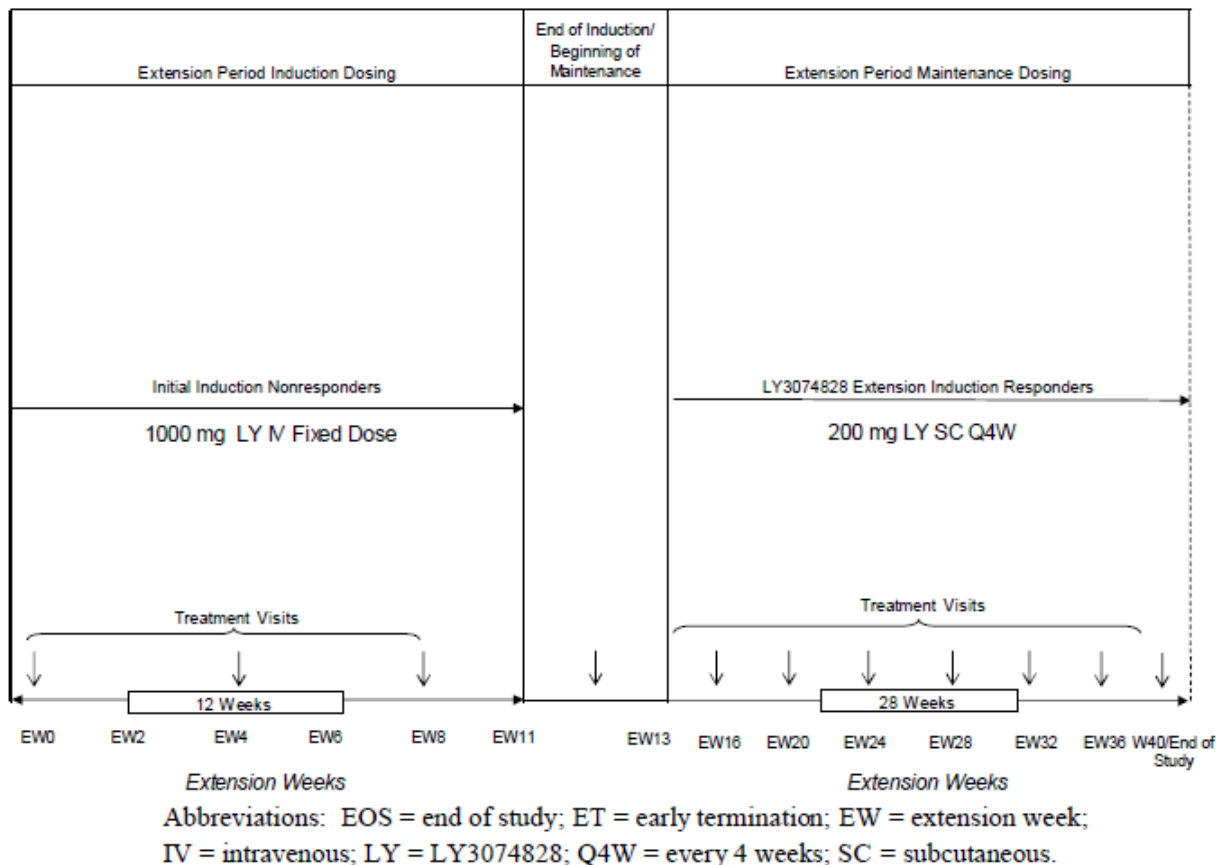
5.2. Determination of Sample Size

Approximately 240 patients will be randomized to 1 of 4 double-blind treatment regimens to evaluate the primary endpoint. Patients will be randomized to placebo, 50 mg LY3074828 Therapeutic Drug Monitoring (TDM), 200 mg LY3074828 TDM, or the 600 mg LY3074828 fixed-dose treatment group in a 1:1:1:1 ratio (60 patients per arm). The randomized patients will be comprised of approximately one-third biologic-naive patients (~20 per arm) and approximately two thirds patients who have been exposed previously to at least 1 biologic agent (~40 per arm).

Assuming LY3074828 and placebo clinical remission rates of 30% and 7.5% respectively (a difference of 22.5%) and given 60 patients per treatment arm, each pairwise comparison testing

the superiority of LY3074828 to placebo will have 89% power via chi-square test with a two - sided 0.05 significance level.

5.3. Method of Assignment to Treatment

Patients who meet all criteria for enrollment will be randomized to treatment at the baseline visit. Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive web-response system (IWRS), and then the site will be responsible for administering the treatment to the patients.

Patients will be randomized into the Induction Dosing Period at a 1:1:1:1 ratio to one of four treatment groups, each given as one IV injection:

- 600 mg LY3074828 IV Q4W: A dose of 600 mg given as an IV infusion at Weeks 0, 4, and 8.

- 200 mg LY3074828 IV TDM Q4W: A dose of 200 mg TDM given as an IV infusion at Weeks 0, 4, and 8. Dose adjustments may occur at Week 4 and 8.

- 50 mg LY3074828 IV TDM Q4W: A dose of 50 mg TDM given as an IV infusion at Weeks 0, 4, and 8. Dose adjustments may occur at Week 4 and 8.

- Placebo: A dose of placebo saline given as IV injections at Weeks 0, 4, and 8.

Patients who have clinical responses to LY3074828 (as defined in section 6.11.3.1) at the end of the Induction Dosing Period will be re-randomized to 1 of 2 maintenance LY3074828 treatment arms in the Maintenance Dosing period. This randomization will be stratified according to the patient's Week 12 remission status.

- 200 mg SC LY3074828 Q4W: A dose of 200 mg given as a SC injection at 4 week intervals starting at week 12.

- 200 mg SC LY3074828 Q12W: A dose of 200 mg given as a SC injection at 12 week intervals starting at week 12.

Patients who have clinical responses to placebo (as defined in section 6.11.3.1) at the end of the Induction Dosing Period will continue to receive placebo via SC injection in the Maintenance Dosing period. Placebo responders will not be re-randomized.

Patients who do not have clinical responses to LY3074828 (as defined in section 6.11.3.1) at the end of the Induction Dosing Period will have the option to continue on to the Extension Period- Induction portion of the trial. Patients entering the Extension Period-_Induction portion of the trial will not be re-randomized but will receive_1000 mg LY3074828 given as an IV infusion at extension weeks 0, 4, and 8.

Patients who have clinical responses to LY3074828 (as defined in section 6.11.3.1) at the end of the Extension Period- Induction will continue to the Extension Period- Maintenance portion of the trial. Patients entering the Extension Period- Maintenance portion of the trial will not be re-randomized but will receive_200 mg LY3074828 given as a subcutaneous (SC) injection at 4 week intervals starting at extension week 12.

6.   A Priori Statistical Methods

6.1.   General Considerations

Eli Lilly and Co. will be responsible for conducting the statistical analysis and for the generation, validation and final production of the Efficacy and Health Outcomes tables, figures and listings (TFLs) proposed in this document. Robarts Clinical Trials, a clinical research organization, will be responsible for conducting the statistical analysis and for the generation, validation and final production of all other tables, figures and listings (TFLs) proposed in this document.

Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol and the justification for making the change will be described in the clinical study report. Additional exploratory analyses of the data will be conducted as deemed appropriate.

Continuous data will be summarized in terms of the number of observations, mean, standard deviation (SD), median, minimum, and maximum. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean and median will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of patients providing data at the relevant time point (n), frequency counts and percentages. Percentages will be presented to one decimal place. Percentages will be calculated using n (the number of observations with non-missing values) as the denominator.

Change from baseline will be calculated as the visit value of interest minus the baseline value. If all baseline values are missing for a particular variable, then the change from baseline and percent change (or percent improvement) from baseline will not be calculated.

All confidence intervals (CIs) and statistical tests will be two-sided unless otherwise specified. P-values which are greater than or equal to 0.001, and less than or equal to 0.999, will be presented to three decimal places. All other p-values which are less than 0.001 will be presented as '<0.001', while p-values greater than 0.999 will be presented as '>0.999'. CIs will be presented to one more decimal place than the raw data.

Age, sex and race will be reported on all by-patient listings unless otherwise specified. Sex will be abbreviated as follows: female (F) and male (M). Race will be abbreviated as follows: American Indian or Alaska Native (AI), Asian (AS), Black or African American (BL), Native Hawaiian or other Pacific Islander (NH), White (WH) and Multiple (MU).

There are five study periods defined in section 7.1 of the protocol:

- Screening Period
- Induction Dosing Period

- Maintenance Dosing Period

- Extension Period- Induction

- Extension Period- Maintenance

- Follow Up Period

Table 1 gives the treatment groups to be displayed for each treatment period and analysis population post implementation of protocol amendment (a). NOTE: "LY3074828" will be included in treatment group names in the statistical reports where possible.

Table 1 Treatment Groups for Each Treatment Period and Analysis Population

Treatment Period	Analysis Population	Treatment Groups	Abbreviation	Between Group comparison (when applicable)	Overall Comparison(when applicable)
Induction Dosing Period	ITT, Safety	600 mg LY3074828 IV 200 mg LY3074828 TDM IV 50 mg LY3074828 TDM IV LY3074828 Total Pbo IV	600 mg IV 200 mg TDM IV 50 mg TDM IV LY Total Pbo IV	600 mg IV vs Pbo IV 200 mg TDM IV vs Pbo IV 50 mg TDM IV vs Pbo IV	600 mg IV 200 mg TDM IV 50 mg TDM IV Pbo IV
Maintenance Dosing Period	Induction Responders	200 SC LY3074828 Q4W 200 SC LY3074828 Q12W LY3074828 Total Pbo SC	200 SC Q4W 200 SC Q12W LY Total Pbo SC	200 SC Q4W vs 200 SC Q12W	No Overall Comparison
Extension Period Induction	All enrolled in Extension Period- Induction	600 mg LY3074828 IV 1000 mg LY3074828 IV Total	600 IV 1000 IV Total	600 mg IV vs 1000 mg IV	No Overall Comparison
Extension Period Maintenance	All enrolled in Extension Period- Maintenance	200 mg LY3074828 SC Q4W	200 SC Q4W	No Between-Group Comparison	No Overall Comparison

Note- Patients who received 600mg IV Extension dosing prior to the protocol amendment will be summarized as 600mg IV. Patients who received 1000mg IV Extension dosing after to the protocol amendment will be summarized as 1000mg IV.

Safety summaries will only compare LY Total to Pbo in Induction.

All summary tables will be presented by treatment group. Where applicable, data will be summarized by treatment group by study visit. Summaries and analyses will be presented

separately for the Induction Dosing Period and the Maintenance Dosing Period. Summaries and analyses will be also presented for the Extension Period.

By-patient listings will be sorted by treatment allocation, site and visit. Numeric data will be listed to the same number of decimal places as recorded on the eCRF. For derived variables, the number of decimal places will be listed as specified. The decimal points will be aligned in the data listings. For the tables and listings in the appendices, "Page x of X" will appear on each page.

By-patient listing will be generated for each SDTM dataset. If necessary, separate listing for derived variables (that are derived from the SDTM datasets) will be generated. Wherever applicable, listings will include the period of the study during which the event or measurement occurred.

Efficacy analyses will be conducted on the Intent-to-Treat (ITT) population. All safety analyses will be conducted on the Safety Population. These populations are defined in section 6.2.

Patients will be analyzed according to the treatment to which they were assigned regardless of any errors of dosing.

All tests of treatment effects will be conducted at a 2-sided alpha level of 0.05 unless otherwise stated. Unless otherwise specified, no multiplicity adjustment will be considered for planned analyses.

For all efficacy and health outcomes analyses, baseline is defined as the last non-missing assessment recorded on or prior to the date of first study drug injection at Week 0 (Visit 2). However, placebo patients who enter the Extension Period will have their baseline reset. Their new baseline will be defined as the last non-missing assessment on or prior to the date of first injection of study treatment at Week EW0 (Extension Visit 1).

Unless otherwise specified, variables will be analyzed in the original scale on which they are measured. However, if it is deemed more statistically appropriate, a transformation, such as to the logarithmic scale, may be applied before analysis. The parametric approach will be employed by default for statistical analysis except when nonparametric analysis, such as by a rank-based method, is assessed to be more fitting.

The primary analysis of categorical (binary) efficacy and health outcome variables will use a logistic regression analysis with treatment, geographic region, and prior biologic experience (prior biologic experience vs prior biologic naïve), and visit (when appropriate) in the model. The odds ratio, the corresponding 95% CIs and p-value, as well as the treatment differences (where appropriate) and the corresponding 95% CIs, will be reported. Secondary analysis will be conducted using an exact Cochran-Mantel-Haenszel test that stratifies by (i.e., controls for) previous biologic use. A Fisher's exact test may be utilized if necessary.

The primary analyses for all continuous efficacy and health outcome variables will be based on the MMRM analysis method. The MMRM analyses will be conducted using a restricted maximum likelihood (REML)-based repeated measures approach. When MMRM is used, the model will include treatment, visit, geographic region, prior biologic experience (prior biologic

experience vs prior biologic naïve), treatment-by-visit interaction, as well as the continuous, fixed covariates of baseline value and baseline value-by-visit interaction. The covariance structure to model the within-patient errors will be unstructured. If the unstructured covariance matrix results in a lack of convergence, the heterogeneous Toeplitz covariance structure, followed by the heterogeneous autoregressive covariance structure, followed by the compound symmetry structure will be used. The first structure to yield convergence will be used for inference. The Newton-Raphson with ridging optimization technique will be used to aid with convergence. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom and adjust standard errors. Type III tests for the LS means will be used for the statistical comparison; the 95% CI will also be reported. Treatment group comparisons with placebo at Week 12 (Visit 7) and all other appropriate time points will be tested.

Secondary treatment comparisons of continuous efficacy variables at Week 12 and 52 will be made using ANCOVA with treatment, geographic region, prior biologic experience (prior biologic experience vs prior biologic naïve), and baseline value in the model. Type III sums of squares for the least-squares (LS) means will be used for the statistical comparison; the 95% confidence interval will also be reported.

Not all displays and analyses described in this SAP will necessarily be included in the CSR. Not all displays will necessarily be created as a "static" display. Some displays may be incorporated as interactive display tools instead of or in addition to a static display. Any display described in this SAP and not provided in CSR would be available upon request.

6.2. Analysis Populations

The patient populations to be analyzed for this study are defined in the table presented below:

Table 2   Analysis Populations

Population	Definition
Screening	All patients who signed informed consent and participated in the screening period (Visit 1). This population will be used for the presentation of patient disposition.
Intent-to-Treat (ITT) Population	Efficacy and health outcome analyses will be conducted on the intent-to-treat population (ITT). The ITT population is defined as all randomized patients, even if the patient does not take the assigned treatment, does not receive the correct treatment, or otherwise does not follow the protocol. Patients will be analyzed according to the treatment to which they were assigned.
Safety Population	Safety analyses will be conducted on the Safety Population, defined as all randomized patients who receive at least one dose of study treatment. Patients will be analyzed according to the treatment to which they were assigned.

Induction Responders	A subset of the ITT population and will include those patients who are re-randomized to one of the two maintenance dosing period LY arms.
All enrolled in Extension Period	A subset of the ITT population and will include only the patients who participated in the Extension Period.

Analyses of the maintenance dosing period and the extension periods will be based on the corresponding enrolled population as described in the table above.

The Schedule of Events outlined in the protocol (attachment 1) specifies the allowable windows for assessments. A sensitivity analysis of the primary analysis may be conducted to investigate the exclusion of patients /assessments deemed to be major deviations from the protocol.

A Per-Protocol population may be evaluated in this study and patients to be excluded from that population will be identified through protocol deviations. All protocol exemptions and violations will be listed as described in section 9.10 of the protocol.

All statistical analyses will be performed using SAS software Version 9.2 (or a higher version). Visualizations tools such as Spotfire may also be used for dynamic displays.

6.3.   Adjustments for Covariates

The randomization at the beginning of the Induction Period is stratified by previous biologic therapy: previous exposure to biologic therapy for treatment of UC, biologic naïve to biologic therapy for treatment of UC. Unless otherwise specified, all efficacy and health outcome analyses will include the stratification factors in the model.

The re-randomization at the beginning of the Maintenance Period of patients who took LY3074828 during the Induction Period and were responders at Week 12 (Visit 7) is stratified by the Week 12 level of clinical response: Clinical Remission or Clinical Response. Unless otherwise specified, all efficacy and health outcomes analyses using the Maintenance Dosing Period Primary population will include the re-randomization factor of Week 12 level of response in the model.

Unless otherwise specified, analyses will be performed utilizing the methodology and covariates described in section 6.1.

6.4.   Handling of Dropouts or Missing Data

Details of the handling of dropouts or missing data are presented in the following sections.

Details of the procedure for handling missing data for the IBDQ are provided in Appendix B.

Details of handling missing partial end dates for prior medications for important programmable deviations are provided in section 6.16 Protocol Violations.

6.4.1. Non-Responder Imputation (NRI)

The primary outcome is the proportion of patients with clinical remission at 12 weeks. For this and other categorical efficacy endpoints (for example, clinical remission (week 52), clinical response (week 12, 52), and endoscopic remission (week 12, 52)), non-responder imputation (NRI) will be used for missing clinical assessment values. Specifically, all patients who discontinue from the study at any time prior to week 12 for any reason or fail to have an adequate week 12 efficacy assessment will be considered a non-responder at week 12. Patients who discontinue from the study for any reason at any time prior to week 52 after having enrolled into the Maintenance Period or fail to have an adequate week 52 efficacy assessment, will be considered a non-responder at week 52.

The NRI may be applied at any time point specified for analysis.

6.4.2. Last Observation Carried Forward

An LOCF analysis will be performed on PGI-I.

This is both a sensitivity analysis, and an analysis for regulatory agencies that prefer this approach. This approach is identical to the mBOCF approach, with one exception: for patients discontinuing investigational product due to an AE, the last non-missing post-baseline observation before discontinuation will be carried forward to the corresponding endpoint for evaluation. Randomized patients without at least one post-baseline observation will not be included for evaluation.

6.4.3. Modified Baseline Observation Carried Forward

An mBOCF analysis will be performed on key continuous efficacy and the PGI-S and the IBDQ.

This is both a sensitivity analysis, and an analysis for regulatory agencies that prefer this approach. For patients discontinuing investigational product due to an AE, the baseline observation will be carried forward to the corresponding primary endpoint for evaluation. For patients discontinuing investigational product for any other reason, the last non-missing post-baseline observation before discontinuation will be carried forward to the corresponding primary endpoint for evaluation. Randomized patients without at least one post-baseline observation will not be included for evaluation with the exception of patients discontinuing study treatment due to an AE.

6.4.4. Missing Data Imputation for Adverse Event, Concomitant Medication Dates, and Laboratory Values

If a medication date or AE date is completely or partially missing, the following imputation rules should be utilized in the analysis unless otherwise stated:

- For the start date:
    - If year, month, and day are missing then use the earlier of the patient's first visit date or the consent date.
    - If either month or month and day are missing then use January 1.

  o If only day is missing, impute the first day of the month.

- For the start time:

  o Impute as 23:59.

- For the end date:

  o If year, month, and day are missing then use the patient's last visit date.

  o If either month or month and day are missing then use December 31.

  o If only day is missing then use the last day of the month.

  o The imputed date will not be beyond the patient's last visit date.

If there is any doubt for the start and end date/times for AEs, the event will be flagged as treatment-emergent. For medications, the medication will be flagged as concomitant.

Laboratory parameters will be summarized based on the available data at each time point and no imputation is proposed in this document.

6.5. Multicenter Studies

This study will be conducted by multiple investigators at multiple sites internationally. Randomization to treatment groups will not be stratified by country. However, the countries will be categorized into geographic regions:

- Geographic Region Subgroups: North America, Asia (Japan) and Rest of World (incl. EU, Australia, and NZ)

Unless otherwise specified, the statistical analysis models will adjust for geographic region. For analysis of primary endpoints, the presence of a treatment-by-geographic region interaction will be tested at a .10% significance level. Treatment group comparisons for the primary outcomes will be presented separately for each geographic region. Where there is evidence of an interaction (p-value $< .0.10$), graphical inspection of descriptive statistics will be used to assess whether the interaction is quantitative (i.e., the treatment effect is consistent in direction but not size of effect) or qualitative (the treatment is beneficial for some but not other regions). In addition, further investigations through additions of baseline covariates of statistical models will be performed in an attempt to explain the interaction in terms of patient characteristics.

6.6. Multiple Comparisons/Multiplicity

The primary endpoint will be tested by using the primary analysis of categorical efficacy method, as described in section 6.1.

A graphical multiple testing procedure [1][2] for the primary objective will be implemented to control the family-wise type I error rate at a 2-sided α level of 0.05. In this particular application, the α levels for the p-values associated with each comparison are computed at each step depending on the outcomes of the preceding significance tests. The graphical approach is a closed testing procedure; hence, it strongly controls the family-wise error rate (Alosh et al. 2014 [2]).

There are 3 comparisons to be tested:

- Comparison 1 – Proportion of 600 mg patients achieving Clinical Remission at 12 weeks compared with placebo

- Comparison 2 – Proportion of 200 mg TDM patients achieving Clinical Remission at 12 weeks compared with placebo

- Comparison 3 – Proportion of 50 mg TDM patients achieving Clinical Remission at 12 weeks compared with placebo

Comparison 1 will first be tested at 2-sided $\alpha = 0.05$. If successful, the test for Comparison 2 will be performed at 2-sided $\alpha = 0.05$ If successful, the test for Comparison 3 will be performed at 2-sided $\alpha = 0.05$. If at any point a comparison fails to meet the significance threshold, the test will stop.

Figure 3   Illustration of Graphical Multiple Testing Procedure with Initial α Allocation and Weights.
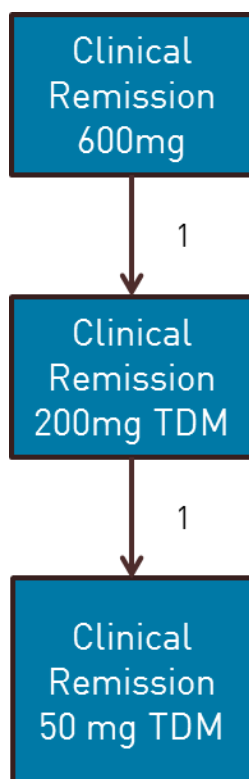
Unless otherwise specified, there will be no adjustment for multiple comparisons for any other analyses.

6.7. Patient Disposition

All patients who discontinue from the study will be identified, and the extent of their participation in the study will be reported. If known, a reason for their discontinuation will be given.

Counts and percentages of patients entered and patients failing screening before randomization will be summarized overall, and counts and percentages of patients randomized, discontinued, and completed will be summarized by treatment group. The reasons for discontinuation from treatment and from the study will be summarized by treatment group.

Summaries of patient disposition will be based on all patients screened in the study. The number of patients in each of the analysis populations will be displayed. The reasons for discontinuation will be summarized. A listing of patients excluded from any population, if any, will be provided along with the reason for the exclusion.

6.8. Patient Characteristics

The patient's year of birth, sex, weight, height, smoking habits, previous biologic treatment, and other demographic characteristics are collected at the screening visit. Age and body mass index will be calculated.

Only the year of birth is collected at screening. For the purpose of age calculation, the month and day of birth will be imputed as July 01, of the year of birth. Age is the computed as follows;

$$Age= (Informed\ Consent\ Date -Date\ of\ Birth\ +1)/365.25.$$

Demographic and baseline characteristics (including age, gender, race and ethnicity) will be summarized for each treatment group.

Certain characteristics, that are collected at baseline or after baseline but not summarized in the demographic summary, will be reported as a listing.

No inferential analysis for the comparability of baseline covariates across treatment groups will be performed.

6.9. Treatment Compliance

Adequate study drug compliance is defined in Section 9.10 of the protocol.

Study treatment administration and compliance will be listed for all entered patients. The number and percentage of patients treatment compliant by week (that is, at each injection time point) will be summarized by treatment group for each treatment period.

No patient will be excluded from the ITT population as a consequence of significant noncompliance.

A listing of the treatment-noncompliant patients will be presented by treatment arm.

6.10. Concomitant Therapy

For the purposes of subgroup identification, medications will be classified into anatomical therapeutic chemical (ATC) drug classes using the latest version of the World Health

Organization (WHO) drug dictionary. Summaries of Concomitant and Prior medication will utilize WHO drug preferred names.

Medication start and stop dates will be compared to the date of first dose of treatment in each treatment period to allow medications to be classified as Concomitant for each treatment period. Concomitant medication use that starts and stops during the baseline/screening period will be included in summaries of the induction period.

Prior medications are those medications that start and stop prior to the date Informed Consent (IC). Concomitant medications are those medications that start after the IC date. Concomitant medications are assigned to the treatment period in which they are actually ongoing. For example, if a patient is receiving concomitant medication during the Induction Dosing Period but has a stop date during the Induction Dosing Period, the same medication would not be listed as a concomitant medication during the latter periods unless patient has a new start date.

6.11. Efficacy Analyses

6.11.1. Primary Outcome and Methodology

Rate of clinical remission at Week 12 is the primary efficacy outcome for this study and will be analyzed using the ITT population.

The primary objective of this study is to test the hypothesis that treatment with LY3074828 is superior to placebo in inducing clinical remission at Week 12 in patients with moderate to severe ulcerative colitis (UC)

Clinical remission is defined as achieving a rectal bleeding Mayo subscore of 0, stool frequency Mayo subscore of 0 or 1 (with 1 point decrease from baseline) and Mayo endoscopic subscore of 0 or 1.

The rates of clinical remission and non-remission will be summarized by treatment group and by the stratification factor, previous biologic use (yes/no). See section 6.1 for details on the methods to be used to test the differences between each active treatment arm and placebo.

Ninety five percent confidence interval for the difference in proportions (LY arm – placebo) for each pairwise comparison will be obtained. If the lower limit of the 95% confidence limit is greater than zero, then that arm will be considered to be superior to placebo.

A multiple testing procedure will be utilized for the primary outcome and is detailed in section 6.6.

See section 6.4.1 for details on NRI.

6.11.2. Additional Analyses of the Primary Outcome

6.11.3. Additional analysis of the clinical remission rate at Week 12 will also be performed as per the methodology described in section 6.1. Analyses of the Secondary Efficacy Outcomes

In the analysis of secondary efficacy outcomes, no adjustments for multiple testing will be performed unless otherwise specified.

6.11.3.1. Secondary Efficacy Outcomes for the Induction Dosing Period

There are two secondary efficacy outcomes for the induction dosing period:

1)   The proportion of patients with clinical response at Week 12.

2)   The proportion of patients with endoscopic remission at Week 12.

The clinical response is defined as achieving at Week 12 a decrease in the 9-point Mayo subscores (comprising the subscores of rectal bleeding, stool frequency and the endoscopic findings) inclusive of ≥ 2 points and $\geq 35\%$ from baseline with either a decrease of rectal bleeding subscore of ≥ 1 or rectal bleeding subscore of 0 or 1.

Endoscopic remission is defined as achieving a Mayo endoscopic score of 0 at Week 12.

Analyses will be performed as per the methodology described in section 6.1.

The ITT population will be used for the secondary efficacy outcome analyses.

6.11.4. Secondary Efficacy Endpoints for the Maintenance Dosing Period

The following table defines the secondary efficacy endpoint evaluations performed for the Maintenance Period:

Table 3   Secondary Efficacy Endpoints for the Maintenance Period

Clinical Remission at Week 52	Achieving the following Mayo subscores at Week 52: a rectal bleeding subscore of 0, stool frequency subscore of 0 or 1 (with 1 point decrease from baseline), and endoscopy subscore of 0 or 1.
Durability of clinical remission at Week 52	Clinical remission at Week 52 for those who had clinical remission at Week 12.
Clinical Response at Week 52	Achieving at Week 52 a decrease in the 9-point Mayo subscores (comprising the subscores of rectal bleeding, stool frequency and the endoscopic findings) inclusive of ≥ 2 points and $\geq 35\%$ from baseline with either

	a decrease of rectal bleeding subscore of >=1 or rectal bleeding subscore of 0 or 1.
Durability of clinical response at Week 52	Clinical response at Week 52 for those who had clinical response at Week 12.
Endoscopic remission at Week 52	Achieving a Mayo endoscopic subscore of 0 at Week 52.
Durability of endoscopic remission at Week 52	Endoscopic remission at Week 52 for those who had endoscopic remission at week 12.

Analyses will be performed as per the methodology described in section 6.1.

6.11.5. Exploratory Efficacy Endpoints

There are three exploratory efficacy endpoints:

1. To assess the change from baseline in Ulcerative Colitis Endoscopic Index of Severity (UCEIS) [4] (completed by central reader only) scores at Weeks 12 and 52.

2. To assess the change from baseline in the partial Mayo score (PGA + Stool Frequency + Rectal Bleeding) at various times during induction and maintenance.

3. To explore the effect of additional (unblinded) dosing on clinical response in patients who do not meet responder criteria at Week 12.

Summary tables for the change from baseline in UCEIS score at weeks 12 and 52 will be presented by treatment groups. The mean and 95% CIs will be presented. The change from baseline will be presented by treatment groups graphically.

Summary tables for the partial Mayo score at assessed time points and the change from the baseline assessments will be presented by treatment groups. The change from baseline will be presented by treatment groups graphically. Subscores will also be summarized separately.

Summary tables of clinical response, clinical remission and endoscopic remission for the patients who were originally randomized to LY3074828 and entered the Extension Period will be generated. Clinical efficacy findings at assessed time points and the change from the baseline assessments will be presented by original Induction Period treatment groups.

In addition to the protocol defined endpoints, additional sensitivity analyses may be performed if deemed appropriate. Additional analyses include:

- Full Mayo Endpoints:
  - Full Mayo Clinical Remission defined as a complete 12 point mayo score of <=2 points and no individual subscore >1 point (Evaluated at week 12, 52)
  - Full Mayo Clinical Response defined as a reduction in complete 12 point mayo score of >=3 points and >= 30% from baseline with an accompanying decrease in

Rectal Bleeding Score of >=1 point OR an Rectal Bleeding Score of <=1 (Evaluated at week 12, 52)

- Partial Mayo Endpoints:
  - Partial Mayo Response (9 point, no Endoscopy): defined as a decrease of >=3 points from baseline
  - Alternate Partial Mayo Response 1 (9 point, no Endoscopy): defined as a decrease of >=2 points and >=25% decrease from baseline with an accompanying decrease in Rectal Bleeding Score of >=1 point OR an Rectal Bleeding Score of <=1
  - Alternate Partial Mayo Response 2 (9 point, no Endoscopy) : defined as a decrease of >=2 points and >=30% decrease from baseline with an accompanying decrease in Rectal Bleeding Score of >=1 point OR an Rectal Bleeding Score of <=1
- Symptomatic Endpoints:
  - Symptomatic Remission (RBS, SF): defined as a SF score of 0 or 1, and a RB score of 0
  - Symptomatic Response (RBS, SF): defined as a decrease of 2 points or greater from baseline
  - Change from baseline in Symptomatic Mayo (SF + RBS) Subscore
- Endoscopic endpoints:
  - Endoscopic Response (evaluated at week 12, 52): defined as at least a 1 point decrease in the endoscopic findings subscore from baseline
  - Endoscopic Healing (evaluated at week 12, 52): defined as achieving an endoscopic findings subscore of 0 or 1
- Time to Event Analyses:
  - Time to Partial Mayo Response (defined below)
  - Time to Symptomatic Response (defined below)
  - Duration of Response in Maintenance (defined below)
- Other Endpoints:
  - Alternate Modified Mayo Remission (9 point, no PGA): a rectal bleeding subscore of 0, stool frequency subscore of 0 or 1 (and no worse than baseline), and an endoscopy subscore of 0 or 1.
  - Alternate Modified Mayo Response (9 point, no PGA): defined as a Modified Mayo score decrease of >=2 points and >=30% decrease from baseline with an accompanying decrease in Rectal Bleeding Score of >=1 point OR an Rectal Bleeding Score of <=1

Partial and Symptomatic analyses will be performed at every visit where collected. Analysis on endpoints including an endoscopic component will only be performed at post baseline visits where an endoscopy was performed (Week 12, Week E12, and Week 52).

Time to Response

Time to response for a given scoring method is defined (in weeks) as:

(First study day when response status is changed to "Yes" – Randomization date + 1) / 7

If a patient has not experienced response by completion or early discontinuation of induction period, the patient will be censored at the date of their last visit during the induction period.

For placebo patients continuing to the extension period, their time to response is defined as:

(First study day when response status is changed to "Yes" – Extension Period Randomization date + 1) / 7

Time to response will be evaluated during the induction and extension periods only.

Duration of Response in Maintenance

The duration of response in maintenance is defined (in weeks) as:

(study day of discontinuation or becoming eligible for rescue - Study day of randomization into Maintenance (or first dose date in Extension Maintenance) +1) / 7

Patients who discontinue early (without loss of response) or complete the study period will be censored. Patients who have a reason for study discontinuation of lack of efficacy or Adverse event will be considered as events.

Duration of response in maintenance will be evaluated during the maintenance and extension maintenance periods only.

Histologic Remission

For the exploratory definitions of histologic remission, analyses will include the presentation of data using spotfire or any other custom analyses deemed appropriate. These presentations are for exploratory purposes only and will be presented in a manner analogous to the proposed histologic remission, described in section 6.14.9.

6.12. Health Outcomes/Quality-of-Life Analyses

6.12.1. Secondary Health Outcomes Endpoints

There are four self-administered questionnaires used for measuring health outcome in this trial:

1. PGI-S (Patient's Global Impressions of Severity score)

2. PGI-I (Patient's Global Impressions of Improvement score)

3. IBDQ (Inflammatory Bowel Disease Questionnaire score)

4. SF36 (36-Item Short Form Health Survey score)

Where appropriate, the total scores and sub-totals for individual dimensions collected will be summarized with means and 95% CIs by time point and by treatment group. The summary table will also include the change from baseline scores wherever applicable. The mean raw scores and mean change from baseline scores with corresponding 95% CIs will be presented graphically by treatment group and in a longitudinal fashion.

The ITT population will be used for all health outcome analyses

Mean change from baseline will be summarized by treatment group.

6.12.1.1. SF36 (36-Item Short Form Health Survey score)

Summaries of domain and summary scores will be provided by visit and change from baseline to postbaseline visits.

An ANCOVA model will be used to analyze the change from baseline in the physical component score (PCS) and the mental component score (MCS) through Week 24. The ANCOVA model will include treatment, geographic region, and prior biologic experience (prior biologic experience vs prior biologic naïve), and baseline value (consistent with adjustments of other efficacy analysis) as fixed effects and baseline score as a covariate with change from baseline as the dependent variable. To test the treatment difference between LY3074828 and placebo, the LS mean difference, 95% CI, and p-value will be presented. The LOCF approach will be used to impute missing data.

Analyses will be repeated for the eight health domains (physical functioning, bodily pain, role limitations due to physical problems, role limitations due to emotional problems, general health perceptions, mental health, social function, and vitality).

Summary statistics per treatment group will be tabulated for the transformed scale scores (physical functioning, role limitations / physical health, bodily pain, general health perceptions, vitality, social functioning, role limitations / emotional, mental health) and the two summary scores (PCS and MCS).

Minimum Clinically Important Differences (MCID):

The summary and analyses of the SF-36v2 Acute Score will include the number of patients achieving a MCID, which is defined for each summary score, physical component score (PCS) and the mental component score (MCS), as:

- MCID $\geq$ at least 2.5 or 5-point improvement (decrease) in PCS
- MCID $\geq$ at least 2.5 or 5-point improvement (decrease) in MCS.

The proportion of patients achieving the SF-36v2 score MCID will also be analyzed using a logistic regression model with treatment, geographic region, and prior biologic experience (prior biologic experience vs prior biologic naïve), and visit (when appropriate) as fixed effects. The p-value and 95% CI of the odds ratio from the logistic regression model will be reported. The 95% CI of the MCID rate difference will be calculated using the Newcombe-Wilson method without continuity correction. The NRI method described in Section 6.4.1 will be used to impute missing data.

6.13. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

PK/PD analyses to address secondary and exploratory objectives of this study will be described by Lilly in separate PK/PD analysis plans. Conduct of the PK/PD analyses will be the responsibility of Eli Lilly and Company.

The PK/PD analyses will be initiated up to approximately three weeks before the planned database lock dates. All analyses will be performed by unblinded Lilly PK personnel, who are unblinded to subject treatment assignments to support dose adjustments, as specified in protocol Section 9.5. Results from the PK/PD analyses will not be shared with site or blinded study team personnel prior to the respective database lock.

6.14. Safety Analyses

All safety evaluations will be based upon the Safety Population as defined in Section 6.2.

Any statistical comparisons for safety data will be performed only between the "All LY3074828" treatment group and Placebo. No pairwise comparisons between LY3074828 induction treatment arms will be performed. Pairwise comparison of LY3074828 maintenance treatment arms will be performed (no comparison to placebo). No multiplicity control procedure will be performed for these tests of safety data.

Safety and tolerability will be evaluated in terms of treatment-emergent AEs (TEAEs), AEs leading to permanent discontinuations of study drug, SAEs, deaths, changes in clinical laboratory test results, and vital signs. These measures will be analyzed for each dose group and combined LY3074828 groups, separately for each study part, as well as across study parts for key summaries.

For all quantitative safety analyses, baseline is defined as the last non-missing assessment recorded on or prior to the date of first study drug injection at Week 0 (Visit 2). However, placebo patients who enter the Extension Period will have their baseline reset. Their new baseline will be defined as the last non-missing assessment on or prior to the date of first injection of study treatment at Week EW0 (Extension Visit 1).

6.14.1. Extent of Exposure

6.14.1.1. Duration of Exposure

Duration of exposure to study treatment (defined as time since first injection of study treatment in days) will be summarized by treatment group during the Induction Period, the Maintenance Period, the Extension Period.

Duration of exposure during the Induction Period for the Safety Population will be calculated as:

(Disposition date (for those who have discontinued the Induction Period), OR Maintenance/Extension Start date -1 (for those who have completed the Induction Period), OR Date of last study visit in the Induction Period (for those who are still being treated in the Induction Period) – Date of first injection of study treatment + 1)

Duration of exposure during the Maintenance Period will be calculated as:

(Disposition date (for those who have discontinued the Maintenance Period), OR Date of last study visit in the Maintenance Period (for those who are still being treated in the Maintenance Period) – Date of the Visit 8 injection of study treatment + 1)

Duration of exposure to LY3074828 during the Extension Period will be calculated as:

(Disposition date (for those who have discontinued the Extension Period), OR Date of last study visit in the Extension Period (for those who are still being treated in the Extension Period) – Date of the Extension Visit 1 injection of study treatment + 1)

Duration of exposure to LY3074828 for the combined Induction Period and Maintenance Period will be calculated as:

(Disposition date (for those who have discontinued the Maintenance Period), OR Date of last study visit in the Maintenance Period (for those who are still being treated in the Maintenance Period)– Date of first injection of LY3074828 + 1)

Duration of exposure to LY3074828 for the combined Induction Period and Extension Period will be calculated as:

(Disposition date (for those who have discontinued the Extension Period), OR Date of last study visit in the Extension Period (for those who are still being treated in the Extension Period)– Date of first injection of LY3074828 + 1)

The date of first injection of LY3074828 is defined in Table 4.

Table 4   First Injection of LY3074828 by Treatment Group

	Visit of first LY3074828
All patients randomized to LY3074828 (regardless of Week 12 response status)	Week 0 (Visit 2)
Week 12 Inadequate Responders originally randomized to placebo and assigned to LY3074828 at Extension Visit 1	Extension Week 0 (Visit EV1)

Descriptive statistics will be provided for patient days of exposure and the frequency of patients falling into the following different exposure ranges (that is, only the exposure ranges that fall within the treatment period will be presented) will be summarized:

- >0, ≥7 days, ≥14 days, ≥30 days, ≥60 days, ≥90 days, ≥120 days, ≥183 days, ≥365 days, ≥548 days.

- >0 to <7 days, ≥7 to <14 days, ≥14 to <30 days, ≥30 to <60 days, ≥60 to <90 days, ≥90 to <120 days, ≥120 to <183 days, ≥183 to <365 days, ≥365 days to <548 days, and ≥548 days

Additional exposure ranges may be considered if necessary.

A by-patient listing of exposure duration will be provided.

Patients who had dose modification will be grouped under the randomized treatment arm and will not be grouped by the modified dose amount.

No inferential analysis for comparison between treatment arms will be performed.

6.14.1.2. Total Dose of active study treatment

The total number of active injections of LY3074828 will be summarized by LY3074828 treatment group.

Mean and median total dose will be reported for all the treatment groups.

Total dose of active study treatment during the Induction Period, the Maintenance Period, the Extension Period, the combined Induction Period and Maintenance Period and the combined Induction Period and Extension Period will be summarized by active treatment group using descriptive statistics. The total dose (mg) will be calculated as:

Total Dose = Sum over all injections received of [dose of LY3074828 (mg) prescribed per injection]

The injections received during the treatment periods will be identified using the response to the question "Was study drug administered to patient?" on the Study administration eCRF page. The total dose of LY3074828 will then be calculated for each treatment group based on the injection schedule using Attachments 1 and 2 in the study protocol.

Patients who had dose modification will be grouped under the randomized treatment arm and will not be grouped by the modified dose amount.

No inferential analysis for comparison between treatment arms will be performed.

6.14.2. Adverse Events

A treatment-emergent adverse event (TEAE) is defined as an event that first occurred or worsened in severity after baseline. The MedDRA Lowest Level Term (LLT) will be used in the treatment-emergent computation. The maximum severity for each LLT during the screening period will be used as baseline. The treatment period will be included as post-baseline for the analysis.

In an overview table, the number and percentage of patients who experienced a TEAE, serious adverse event, adverse event related to study drug, died due to an adverse event, or discontinued from the study due to an adverse event will be summarized by treatment. These summary tables will additionally be provided for the Induction, Maintenance and Extension Periods.

The percentages of patients with TEAEs will be summarized by treatment using MedDRA Preferred Term nested within System Organ Class. Events will be ordered by decreasing frequency within System Organ Class. As an additional table, the percentages of patients with TEAEs will be summarized by treatment using MedDRA Preferred Term (without regard to System Organ Class). Events will be ordered by decreasing frequency.

Adverse events considered possibly related to study drug by the investigator will be displayed using MedDRA Preferred Term nested within System Organ Class. Events will be ordered by decreasing frequency within System Organ Class.

Adverse events that start during the screening period will be summarized in listings..

Subgroup analysis by Prior Biologic exposure will be performed for the summaries described above. Prior Biologic use status for the purposes of subpopulation identification will be based on the 'Prior Biologic Y/N?' eCRF question.

Brief narratives of all patients experiencing a death, a serious adverse event, or an adverse event leading to study drug discontinuation will be provided.

6.14.2.1. Common Adverse Events

The percentages of patients with TEAEs will be summarized by treatment using MedDRA Preferred Term nested within System Organ Class for the common TEAEs (occurred in >=10% before rounding of treated patients). Events will be ordered by decreasing frequency within System Organ Class.

The percentages of patients with TEAEs by maximum severity will be summarized by treatment using MedDRA Preferred Term nested within System Organ Class for TEAEs. For each subject and TEAE, the maximum severity for the MedDRA level being displayed (Preferred Term, High Level Term, or System Organ Class) is the maximum post-baseline severity observed from all associated LLTs mapping to that MedDRA level. These summary tables will additionally be provided for the Induction, Maintenance and Extension Periods.

In the event differential dropout rates are seen or to further investigate events of interest, summary tables comparing exposure-adjusted incidence rate (that is, person-time-adjusted incidence rates) over the entire time period will be generated for the Induction and Maintenance Periods.

6.14.2.2. Hypersensitivity Reactions

Two main analyses are performed in support of assessment of potential immediate hypersensitivity, including anaphylaxis and infusion related reactions (IRR), and well as potential non-immediate hypersensitivity.

Time Period A, of potential immediate hypersensitivity, includes all TEAE occurring on the day of study drug administration.

Time Period B, of potential non-immediate hypersensitivity, includes all TEAE occurring strictly after the day of study drug administration (but prior to subsequent drug administration).

Analyses of both time periods use the current standard Medical Dictionary for Regulatory Activities (MedDRA) Standardised Medical Queries (SMQs) to search for relevant events. TEAE are characterized as follows:

- Anaphylactic reaction SMQ (20000021; narrow, algorithm per MSSO SMQ guide, and broad)

- Hypersensitivity SMQ (20000214; narrow and broad)
- Angioedema SMQ (20000024; narrow and broad)
- Preferred term (PT) of Infusion related reaction (10051792)

The number and percentage of patients who experienced a TEAE for the following will be analyzed for each of the two time periods:

- Any narrow or algorithmic term from any one of the 3 SMQs indicated above (that is, combined search across narrow and algorithmic portions of all 3 SMQs)
- Any narrow scope term within each SMQ, separately (that is, narrow SMQ search)
- Any term within each SMQ, separately (that is, broad SMQ search)

Within query, individual PTs that satisfy the queries will be summarized. For Infusion related reaction (PT), the individual LLTs will be summarized.

The Anaphylactic reaction SMQ algorithm will be run only for Time Period A. The SMQ defines a category (A, B, C, D) for each SMQ PT. All Narrow terms have category A, and the occurrence of a Narrow term satisfies the algorithm. Additionally, a pair of PTs *from the same drug administration* satisfies the algorithm if the two events are from different categories (*i.e.*, B&C, or B&D, or C&D). Both contributing events must begin within Time Period A. Tables will summarize (the number of patients experiencing) each PT that contributes to such an algorithmic pair, and include such terms in the combined narrow and algorithmic search. Broad events that do not contribute to the algorithm will be summarized in a distinct portion of the table.

For Time Period A only, the number and percentage of each PT that is not in any of the 3 SMQs will be summarized overall and by individual PT. Only PTs that occur in at least 3 patients receiving LY3074828 will be displayed in this portion of the table.

The PT and LLT are listed for summary in decreasing order of incidence for LY-treated patients. Note that an individual patient may contribute multiple events. Also, a single event may satisfy multiple SMQ, in which case the event contributes to every applicable SMQ.

The analyses of this section will be performed for Induction Period, for Induction and Maintenance Periods combined, and for Induction and Extension Periods combined.

6.14.2.3. Analysis of Injection Site Reactions

A summary will be provided, by treatment group, of the number of patients with reported events that map to any one of the following:

- MedDRA High Level Term (HLT) of Injection site reaction

- HLT of Administration site reaction

- HLT of Infusion site reaction.

The number and percentage of patients who experienced a TEAE for the following will be analyzed:

- Any term from either of the HLTs indicated above (that is, a combined query of these HLTs)
- Any term within each HLT separately

Within HLT, individual PTs that satisfy the queries will be summarized.

The PT will be listed for summary with each category in decreasing order of incidence for LY-treated patients.

The analyses of this section will be performed for Induction Period, for Induction and Maintenance Periods combined, and for Induction and Extension Periods combined.

6.14.3. Deaths, Serious Adverse Events, and Other Significant Adverse Events

The number and percentage of patients who experienced a serious adverse event during the treatment period will be summarized by treatment using MedDRA Preferred Term nested within System Organ Class. Events will be ordered by decreasing frequency within System Organ Class.

These summary tables will be generated for the study overall and additionally will be generated for the Induction, Maintenance and Extension Periods.

In the event differential dropout rates are seen, summary tables comparing exposure-adjusted incidence rate (that is, person-time-adjusted incidence rates) over the entire time period will be generated for the Induction and Maintenance Periods for applicable safety evaluations of interest.

The number and percentage of subjects who permanently discontinued from study treatment due to an adverse event (including adverse events that led to death) during the treatment period will be summarized by treatment using MedDRA Preferred Term nested within System Organ Class. Events will be ordered by decreasing frequency within System Organ Class.

A listing of deaths will be provided. All deaths will be included, regardless of the Investigator's or the Sponsor's judgment about causality, including (1) any deaths occurring during participation of the study, (2) any deaths occurring after a patient leaves (i.e., discontinued from the study or completes the study the study if the death is the result of a process initiated during the study, regardless of when it actually occurs. Each listing will include investigator ID, patient ID, treatment group, baseline age, sex, associated AE, whether or not the death and Lilly's assessment of whether the process leading to death (NOT the death itself) began:

- "On study": during study or within the 24-hour day after day of last dose in study.

- "Shortly after study": from the end of the "On Study" time period to 4 weeks after last dose in study or

- "Long after study": more than 4 weeks (or longer, as specified in "Shortly After Study") after the last dose in study

6.14.4. Clinical Laboratory Evaluation

Laboratory evaluations will be summarized and analyzed for the following periods:

- Induction Period,

- Maintenance Period,

- Extension Period (including Extension Induction and Extension Maintenance)

Laboratory tests include all planned analytes as defined in the protocol, excluding those collected in a reflex manner (only collected under certain circumstances).

Values at each visit (starting at randomization) and change from baseline to each visit for laboratory tests will be displayed in box plots (notched box for each treatment with outliers displayed) for patients who have both a baseline and a result for the specified visit. Individual measurements outside of reference limits (defined by the performing laboratory) will also be displayed using distinct symbols overlaying the box plot. Baseline will be the last non-missing observation in the screening period. Original-scale data will be used for the display. Unplanned measurements will be excluded. Displays using both SI and U.S. conventional units will be provided (when different). The following summary statistics will be included in a table below the box plot: N, mean, standard deviation, minimum, Q1, median, Q3, and maximum. Box plots will be used to evaluate trends over time and to assess a potential impact of outliers on central tendency summaries.

Change from baseline to last observation will also be summarized within the box plot of changes (rightmost column) for patients who have both baseline and at least one post-baseline result. Baseline will be the last non-missing observation in the screening period. The last non-missing observation in the treatment period will be used as the last observation. Original-scale data will be used. Unplanned measurements will be excluded. A p-value will be included in the summary statistics at the bottom of the box plot for this assessment, using an ANCOVA model containing terms for treatment and the continuous covariate of baseline measurement.

For quantitative laboratory analyte measurements, 3-panel displays that include a scatterplot, a shift table, and a shift to high/low table will be created. Specifically, for each measurement, both a 3-panel display assessing low values and a 3-panel display assessing high values will be created.

In the 3-panel display to assess low values, the scatterplot will plot the minimum value during the screening period versus the minimum value during the treatment period. Lines indicating the reference limits are included. In cases where limits vary across demographic characteristics, lines indicating the most common limit will be displayed. The shift table will include the number and percentage of patients within each baseline category (minimum value is low, normal, high, or missing) versus each treatment category (minimum value is low, normal, or high) by treatment. Patients with at least one result in the treatment period will be included in the shift table. The shift from normal/high to low table will include the number and percentage of patients by treatment whose minimum baseline result is normal or high and whose minimum treatment result

is low. Patients whose minimum baseline result is normal or high and have at least one result during the treatment period are included. The Fisher's exact test will be used to compare percentages of patients who shift from normal/high to low between treatments.

The 3-panel display to assess high values will be created similarly. The scatterplot will plot the maximum value during the screening period versus the maximum value during the treatment period. The shift table will include the number and percentage of patients within each baseline category (maximum value is low, normal, high, or missing) versus each treatment category (maximum value is low, normal, or high) by treatment. The shift from normal/high to low table will include the number and percentage of patients by treatment whose maximum baseline result is normal or low and whose maximum treatment result is high. Patients whose maximum baseline result is normal or low and have at least one result during the treatment period are included.

For laboratory analyte measurements collected qualitatively, a listing of abnormal findings will be created. The listing will include patient ID, treatment group, laboratory collection date, analyte name, and analyte finding.

6.14.4.1. Hepatic Laboratory Examinations

Analyses for laboratory analyte measurements are described in Section 5.3. This section describes additional analyses for the topic.

To further evaluate potential hepatotoxicity, an eDISH (Evaluation of Drug-Induced Serious Hepatotoxicity) plot will be created. Each subject with at least one post-baseline ALT and total bilirubin contributes one point to the plot. The maximum ALT measurement and the maximum total bilirubin measurement during the treatment period are used. The measurements do not need to be taken at the same blood draw.

The percentages of subjects with the following elevations in hepatic laboratory tests at any time will be summarized between treatment groups:

- The percentages of patients with a alanine aminotransferase (ALT) measurement greater than or equal to 3 times (3X), 5 times (5X), and 10 times (10X) the performing laboratory upper limit of normal (ULN) during the treatment period will be summarized for all patients with a post-baseline value and for subsets based on various levels of baseline value.

  - The analysis of 3X ULN will contain 4 subsets:
    - patients whose non-missing maximum baseline value is less than or equal to 1X ULN,
    - patients whose maximum baseline is greater than 1X ULN but less than 3X ULN,
    - patients whose maximum baseline value is greater than or equal 3X ULN, and
    - patients whose baseline values are missing.

- o The analysis of 5X ULN will contain 5 subsets:
  - patients whose non-missing maximum baseline value is less than or equal to 1X ULN,
  - patients whose maximum baseline is greater than 1X ULN but less than 3X ULN,
  - patients whose maximum baseline is greater than or equal to 3X ULN but less than 5X ULN,
  - patients whose maximum baseline value is greater than or equal to 5X ULN, and
  - patients whose baseline values are missing.

- o The analysis of 10X ULN will contain 6 subsets:
  - patients whose non-missing maximum baseline value is less than or equal to 1X ULN,
  - patients whose maximum baseline is greater than 1X ULN but less than 3X ULN,
  - patients whose maximum baseline is greater than or equal to 3X ULN but less than 5X ULN,
  - patients whose maximum baseline is greater than or equal to 5X ULN but less than 10X ULN,
  - patients whose maximum baseline value is greater than or equal to 10X ULN, and
  - patients whose baseline values are missing.

- The percentages of patients with an aspartate transaminase (AST) measurement greater than or equal to 3 times (3X), 5 times (5X), and 10 times (10X) the performing laboratory ULN during the treatment period will be summarized for all patients with a post-baseline value and for subsets based on various levels of baseline, as described above for ALT.

- The percentages of patients with a total bilirubin measurement greater than or equal to 2 times (2X) the performing laboratory ULN during the treatment period will be summarized for all patients with a post-baseline value, and subset into 4 subsets:

  - o patients whose non-missing maximum baseline value is less than or equal to 1X ULN,
  - o patients whose maximum baseline is greater than 1X ULN but less than 2X ULN,
  - o patients whose maximum baseline value is greater than or equal to 2X ULN, and
  - o patients whose baseline values are missing.

Maximum baseline will be the maximum non-missing observation in the screening period. The maximum value will be the maximum non-missing value from the treatment period. Planned and unplanned measurements will be included.

In addition, the percentages of subjects with treatment-emergent potentially drug-related hepatic disorders will be summarized by treatment by using the MedDRA preferred terms contained in any of the following SMQs:

- Broad and narrow terms in the Liver related investigations, signs and symptoms SMQ (20000008)

- Broad and narrow terms in the Cholestasis and jaundice of hepatic origin SMQ (20000009)

- Broad and narrow terms in the Hepatitis non-infections SMQ (20000010)

- Broad and narrow terms in the Hepatic failure, fibrosis and cirrhosis and other liver damage SMQ (20000013)

- Narrow terms in the Liver-related coagulation and bleeding disturbances SMQ (20000015)

The percentage of subjects with any one of the terms will be summarized in addition to the percentages for each MedDRA preferred term. The percentages of subjects with potentially drug-related hepatic disorders that led to permanent study treatment discontinuation will be summarized similarly.

These analyses will be provided for the Induction and Maintenance Periods. Only the summary of ALT (3X, 5X, and 10X), AST (3X, 5X, and 10X), and total bilirubin (2X) at any time will be provided for the Extension period.

Individual graphical patient profiles will be prepared for subjects with an ALT or AST measurement greater than or equal to 5X ULN or with an alkaline phosphatase measurement greater than or equal to 2X ULN. A graphical patient profile will be created for any patient meeting the criteria from the safety population (any phase, any mediation). The graphical patient profile will include demographics, disposition, and a display of study drug exposure, adverse events, medications, and the liver-related measurements over time. The review for these subjects includes an assessment of the proximity of any ALT or AST elevation to any total bilirubin elevation, alkaline phosphatase levels, gamma-glutamyl transpeptidase (GGT) levels, other potential causes, and the temporal association with events such as nausea, vomiting, anorexia, abdominal pain, or fatigue.

A plot of maximum post-baseline ALT vs. maximum post-baseline total bilirubin will be created that includes all patients from the safety population. Each subject with at least one post-baseline ALT and total bilirubin contributes one point to the plot. The maximum ALT measurement divided by upper limit of normal (ULN) and the maximum total bilirubin measurement divided by ULN during the treatment period are used. The measurements do not need to be taken at the

same blood draw. Symbols will be used to indicate treatments (concatenated when multiple treatments are taken).

6.14.5. Vital Signs and Other Physical Characteristics

Vital signs include systolic blood pressure, diastolic blood pressure, pulse, and temperature. Physical characteristics include weight and BMI.

Vital signs and physical characteristics will be summarized and analyzed for the following periods:

- Induction Period
- Maintenance Period
- Extension Period

Vital sign and physical characteristic results will be provided in a by-patient listing.

BMI will be calculated using the following formula.

BMI (kg/m2) = weight (kg) divided by the square of the height (m).

Values at each visit (starting at randomization) and change from baseline to each visit for vital signs and physical characteristics will be displayed in box plots (notched box for each treatment with outliers displayed) for patients who have both a baseline and a result for the specified visit. Individual measurements outside of reference limits will also be displayed using distinct symbols overlaying the box plot. Baseline is defined as the last non-missing assessment recorded on or prior to the first study drug injection at Week 0 (Visit 2). Original-scale data will be used for the display. Unplanned measurements will be excluded. The following summary statistics will be included in a table below the box plot: N, mean, standard deviation, minimum, Q1, median, Q3, and maximum. Box plots will be used to evaluate trends over time and to assess a potential impact of outliers on central tendency summaries.

Change from baseline to last observation will also be summarized within the box plot of changes (rightmost column) for patients who have both baseline and at least one post-baseline result. Baseline is defined as the last non-missing assessment recorded on or prior to the first study drug injection at Week 0 (Visit 2). The last non-missing observation in the treatment period will be used as the last observation. Original-scale data will be used. Unplanned measurements will be excluded. A p-value will be included in the summary statistics at the bottom of the box plot for this assessment, using an ANCOVA model containing terms for treatment and the continuous covariate of baseline measurement. The percentages of patients with treatment-emergent high or low vital signs and physical characteristics results at any time will be summarized. A treatment-emergent high result is defined as a change from a value less than or equal to the high limit at all baseline visits to a value greater than the high limit at any time that meets the specified change criteria during the treatment period. A treatment-emergent low result is defined as a change from a value greater than or equal to the low limit at all baseline visits to a value less than the low limit at any time that meets the specified change criteria during the treatment period. To assess increases, change from the maximum value during the screening period to the maximum value

during the treatment period will be used. To assess decreases, change from the minimum value during the screening period to the minimum value during the treatment period will be used. Tables 5 will be used to define the low and high limits and change thresholds.

Additionally, values at each visit (starting at randomization) and change from baseline to each visit for vital signs and physical characteristics will be displayed in box plots (with outliers displayed) for patients who have both a baseline and a result for the specified visit. Baseline is defined as the last non-missing assessment recorded on or prior to the first study drug injection at Week 0 (Visit 2). Original-scale data will be used for the display. Unplanned measurements will be excluded. Reference limits will not be displayed. The following summary statistics will be included in a table below the box plot: N, mean, standard deviation, minimum, Q1, median, Q3, and maximum. Box plots will be used to evaluate trends over time and to assess a potential impact of outliers on central tendency summaries.

Table 5   Categorical Criteria for Abnormal Treatment-Emergent Blood Pressure and Pulse Measurement, and Categorical Criteria for Weight and Temperature Changes for Adults

Parameter	Low mmHg	High mmHg
Systolic BP (mm Hg) (Supine or sitting – forearm at heart level)	≤ 90 and decrease from baseline ≥ 20	≥ 140 and increase from baseline ≥ 20
Diastolic BP (mm Hg) (Supine or sitting – forearm at heart level)	≤ 50 and decrease from baseline ≥ 10	≥ 90 and increase from baseline ≥ 10
Pulse (bpm) (Supine or sitting)	< 50 and decrease from baseline ≥ 15	> 100 and increase from baseline ≥ 15
Weight (kg) (Consistent clothing and timing in relationship to meals and voiding)	(Loss) decrease $\geq 7\%$	(Gain) increase $\geq 7\%$
Temperature	< 96 degrees F and decrease ≥ 2 degrees F	≥ 101 degrees F and increase ≥ 2 degrees F

6.14.6. Electrocardiograms

Detailed Electrocardiogram results are part of the clinical database for this study. Should an ECG be performed in association with an AE or medical history event, the occurrence of the ECG ("Yes/No") will be provided in a by-patient listing.

If any clinically significant ECG measurement occurs, this will be recorded as an AE. ECG data will not be analyzed.

6.14.7. Biomarkers

Samples for the following biomarkers are collected at various timepoints during the trial:

- CRP

- Faecal Calprotectin

- Serum MMP9

- Serum NGAL

Summary tables will be provided by treatment arms at each timepoint the sample was collected.

A graphical representation for the observed value and the change from baseline over time will be presented by treatment group.

The above summaries may be presented separately for each study period if appropriate.

Correlational analysis to determine the relationship of LY3074828 dose levels to biomarker readings will be performed.

Additional correlational analyses to investigate the biomarker readings as a predictor of efficacy outcomes (Clinical remission/response and endoscopic remission) and health outcomes questionnaire findings may be performed if considered useful in further understanding of the study exploratory objectives.

6.14.8. Immunogenicity

Figure 1 provides an overview of the immunogenicity assay process.

At a high level, an individual sample is potentially examined multiple times, in a hierarchical procedure, to produce a sample Anti-Drug Antibodies (ADA) assay result and potentially a sample neutralizing antibody (NAb) assay result. The cut points used, the drug tolerance of an assay, and the possible values of titers are operating characteristics of the assay.

It can be the case that the presence of high concentrations of LY will affect the measurements of the presence of ADA or NAb, and conversely high levels of ADA or NAb may affect the measurement of LY concentration. Thus an LY drug concentration, assessed from a sample drawn at the same time as the ADA sample, plays a key role in clinical interpretation of a sample when the laboratory result is Not Detected, as shown in Figure 1.

The rest of this section defines the component concepts of Figure 1 in greater detail.

6.14.8.1. Definitions of Sample ADA Status

Table 6   Sample ADA Assay Results

Sample Laboratory Result	Explanation
Detected	ADA are detected and confirmed.
Not Detected	The raw result as reported from the laboratory indicates not detected. The clinical interpretation of such results depends on other factors (see below).
NO TEST, QNS, etc.	Sample exists but was unevaluable by the assay

Table 7   **Sample Clinical ADA Interpretation Results**

Sample Clinical Interpretation	Explanation
ADA Present	ADA assay result is Detected
ADA Not Present	ADA assay result is Not Detected <u>and</u> simultaneous drug concentration is at a level that has been demonstrated to not interfere in the ADA detection method (i.e. drug concentration is below the assay's drug tolerance level). For patients receiving placebo, drug concentration is not assessed and is assumed to be below the assay's drug tolerance level. If drug concentration was planned but is not available for a treatment-period sample, a Not Detected sample will be declared ADA Not Present, on the basis of prior knowledge that the drug tolerance level of the ADA assay is high relative to peak PK levels.
ADA Inconclusive	ADA assay result is Not Detected but drug concentration in the sample is at a level that can cause interference in the ADA detection method.
ADA Missing	ADA sample not drawn, QNS, not tested, etc., causing there to be no laboratory result reported or the result is reported as "no test".

Parallel terminology applies for Neutralizing ADA (NAb) Detected, NAb Not Detected, NAb Present, NAb Not Present, NAb Inconclusive, NAb Missing. ADA and Neutralizing ADA (NAb) are distinct assays and have different assay operating characteristics.

Figure 4   Flow Chart of ADA Sample Assessment with Clinical Interpretation
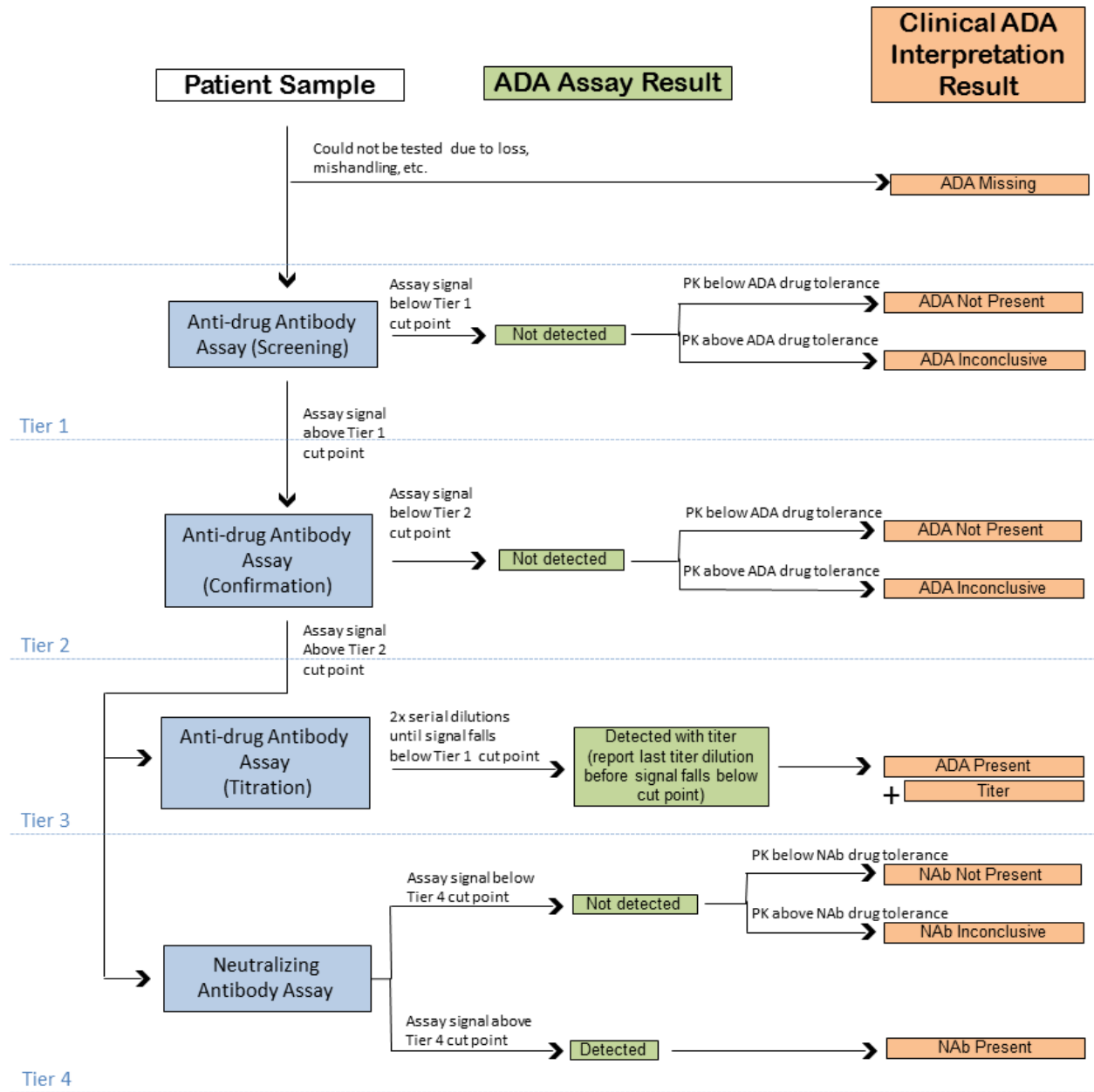
6.14.8.2. Definitions of Immunogenicity Assessment Periods

Immunogenicity Baseline Observations: Screening period for immunogenicity assessment for each patient includes all observations on or prior to first administration of study drug. In instances where multiple baseline observations are collected, to determine patient ADA status the last non-missing immunogenicity assessment prior to first administration of study drug is used to determine treatment-emergent status (see below). In this context, 'missing' includes explicit 'ADA Missing' results, as defined in Table 7.

Immunogenicity Postbaseline Period Observations: Postbaseline period observations for each patient includes all observations after the first administration of study drug.

6.14.8.3. Definitions of Patient ADA Status

Patient evaluable for treatment-emergent ADA: A patient is evaluable for treatment-emergent ADA if the patient has a non-missing baseline ADA result, and at least one non-missing post-baseline ADA result.

Treatment-emergent ADA positive (TE ADA+) patient: A patient who is evaluable for treatment-emergent ADA is treatment-emergent ADA positive (TE ADA+) if either of the following holds:

a. The patient has baseline status of ADA Not Present and at least one post-baseline status of ADA Present with titer $\geq 2*$MRD, where the MRD is the minimum required dilution of the ADA assay.
b. The patient has baseline and post-baseline status of ADA Present, with the post-baseline titer being 2 dilutions (4-fold) greater than the baseline titer. That is, the patient has baseline status of ADA Present, with titer 1:B, and at least one post-baseline status of ADA Present, with titer 1:P, with P/B ≥ 4.

Treatment-emergent ADA Inconclusive patient: A patient who is evaluable for TE ADA is TE ADA Inconclusive if $\geq 20\%$ of the patient's post-baseline samples, drawn pre-dose, are ADA Inconclusive and the patient is not otherwise TE ADA+ (i.e. over all postbaseline visits, the patient does not have a TE ADA+ titer).

Treatment-emergent ADA negative (TE ADA-) patient: A patient who is evaluable for TE ADA is treatment-emergent ADA negative (TE ADA-) when the patient is not TE ADA+ and the patient is not TE-ADA Inconclusive.

6.14.8.4. Analyses to be performed

A listing will be provided of all immuno assessments for those patients who at any time had ADA Present. This includes the laboratory ADA assay result (Detected or Not Detected), LY concentration from a simultaneous PK sample, and the clinical interpretation result that combines these (ADA Present, ADA Not Present, ADA Inconclusive, Missing). When Detected, a titer will be included, and TE ADA+ observations will be flagged. Also included, when the NAb assay was performed, will be the laboratory NAb assay result (Detected or Not Detected)

and the NAb clinical interpretation result (NAb Present, NAb Not Present, NAb Inconclusive, Missing).

In the remainder of this section, "ADA result" will refer to the clinical interpretation result, and similarly for NAb.

The number and proportion of patients who are TE ADA+ will be tabulated by treatment group, where proportions are relative to the number of patients who are TE ADA evaluable, as defined above. The tabulation will include all post-baseline observations and will include the number and proportion of patients with ADA Present at baseline, and also the number and proportion of TE ADA+ patients exhibiting NAb+. This summary will be provided for Induction Period, for Induction and Maintenance Periods combined, and for Induction and Extension Periods combined.

For each TE ADA+ patient, a plot will be constructed of titer values from individual samples over time. Samples that are ADA Not Present or ADA Inconclusive will also be indicated.

A summary will be provided of the number and percentage of LY-treated patients experiencing specific TEAE (see Table 8) by patient TE ADA status (TE ADA+, TE ADA-, TE ADA Inconclusive). The PT will be ordered by decreasing incidence in TE ADA+ status group. For event terms not occurring in any TE ADA+ or TE ADA Inconclusive patient, a summary will be provided of the number of distinct terms reported, but individual PT will not be summarized. This summary will be provided for Induction and Maintenance Periods combined and for Induction and Extension Periods combined.

A listing will be provided of all TEAE alongside ADA data, for any patient who had ADA Present at any time (including baseline) <u>or</u> had any specific TEAE (see Table 8). This listing includes a time course of ADA (clinical interpretation result plus flags for samples meeting TE ADA+ criteria and for NAb+ samples) along with the TEAE.

Table 8   AE for Analysis with ADA / NAb Results

Events satisfying Anaphylaxis SMQ (narrow or broad)
Events satisfying Hypersensitivity SMQ (narrow or broad)
Events satisfying Angioedema SMQ (narrow or broad)
Events mapping to Preferred Term (PT) of Infusion related reaction
Events mapping to High Level Term (HLT) of Injection site reaction
Events mapping to High Level Term (HLT) of Infusion site reaction
Events mapping to High Level Term (HLT) of Administration site reaction

6.14.9. Histopathology

The histopathologic images will be read centrally in a blinded manner by a qualified pathologist and scoring performed using the Geboes Score and Robarts Histopathology Index (RHI) [3].

The Geboes score assigns subjects to grades 0, 1, 2A, 2B, 3, 4 and 5 according to seven histological features: structural (architectural change), chronic inflammatory infiltrate, lamina

propria eosinophils, lamina propria neutrophils, neutrophils in epithelium, crypt destruction and erosion or ulceration [3].

In practice, the highest grade in which there is evidence of disease is assigned. For example if <50% crypts involved (3.2) is checked and Crypt destruction is noted as 'none' (4.0) and Erosion or ulceration is 'No erosion, ulceration ,or granulation tissue' (5.0), the subject will be assigned a score or grade of 3. For scoring, a subject with either abnormalities in lamina propria eosinophils or lamina propria neutrophils will be assigned a score or grade of 2.

The Robarts Histopathology Index is based on the chronic inflammatory infiltrate, lamina propria neutrophils, neutrophils in epithelium, and erosion ulceration components of the Geboes score. As reported in Mosli et al, the RHI is calculated as:

$$RHI = 1 \text{ x chronic inflammatory infiltrate level (4 levels)}$$

$$+ 2 \text{ x lamina propria neutrophils (4 levels)}$$

$$+ 3 \text{ x\_neutrophils in epithelium (4 levels)}$$

$$+ 5 \text{ x erosion or ulceration (4 levels after combining Geboes 5.1 and 5.2).}$$

The total score ranges from 0 (no disease activity) to 33 (severe disease activity).

The following proposed definition for histologic remission will be used for the purpose of analyses and interpretation in this study:

- **Proposed definition: Histologic remission,** defined as Geboes histological subscores of 0 for parameters:
  o 2B (neutrophils in lamina propria), and
  o 3 (neutrophils in epithelium), and
  o 5 (erosion or ulceration)

The following three additional definitions are also presented for exploratory purposes only:

- **Exploratory definition 1: Histologic remission,** defined as Geboes histological subscores of 0 for parameters:
  o 2B (neutrophils in lamina propria), and
  o 3 (neutrophils in epithelium), and
  o 5 (erosion or ulceration)
  and
  o 0 (structural [architectural change]) – no worse than baseline.

- **Exploratory definition 2: Histologic remission,** defined as Geboes histological subscores of 0 for parameters:
  o 2B (neutrophils in lamina propria) and
  o 3 (neutrophils in epithelium), and
  o 4 (crypt destruction), and
  o 5 (erosion or ulceration)

- **Exploratory definition 3: Histologic remission,** defined as Geboes histological subscores of 0 for parameters:
  o 2B (neutrophils in lamina propria)
  o 3 (neutrophils in epithelium), and
  o 4 (crypt destruction), and
  o 5 (erosion or ulceration)
  and
  o 0 (structural [architectural change]) – no worse than baseline.

Summary for the ITT population by treatment arm will be presented for each timepoint a colon biopsy was performed. Also, the change from baseline at Week 12 of the induction dosing period will be presented by treatment arm.

Subgroup summaries by previous biologics exposure, by baseline severity and any other clinically significant baseline stratification may be performed.

Two sample unpaired t-test will be employed to analyze the mean difference at Week 12 and Week 52 between each active treatment arm and placebo for each scoring method. The 95% confidence interval for the mean difference will be obtained. Similar analysis results will be presented for the previous biologics exposure subgroups.

6.15. The proportion of subjects with a Histologic Remission at week 12 and week 52 will be presented and analysed as per the methodology described in section 6.1.Subgroup Analyses

6.15.1. Efficacy Subgroup Analysis

Subgroup analyses will be conducted for the following efficacy assessments:

- Clinical Remission at 12 weeks

Subgroups to be evaluated include the following:

- Patient Demographic Subgroups:

  o Gender: female or male

  o Age group: < 65 years, ≥65 years

  o Race: American Indian/Alaska Native, Asian, Black/African American, Native Hawaiian/Other Pacific Islanders, White, or Multiple Races

  o Ethnicity: Hispanic/Latino, Non-Hispanic/Non-Latino

  o Weight: <100 kg or ≥100 kg

  o Weight: <80 kg, ≥80 and <100 kg, or ≥ 100 kg

- o Body mass index (BMI, kg/m^2) category (underweight (<18.5 kg/m^2); normal (≥ 18.5 and <25 kg/m^2); overweight (≥25 and <30 kg/m^2); obese (≥ 30 and < 40 kg/m^2); or extreme obese (≥ 40 kg/m^2))

- Geographic Region Subgroups:

  - o Geographic region: North America, Asia (Japan) and Rest of World (incl. EU and Australia, NZ)

- Prior/Concomitant UC Therapy:

  - o Prior Biologic therapy vs Biologic Naïve

  - o Concomitant UC Baseline Medication:

    - ▪ Corticosteroid current use (concomitant use of any medications with ATC codes A07EA, H02AB, A01AC) vs none

    - ▪ 5 ASAs current use (concomitant use of any medications with ATC code A07EC) vs none

    - ▪ Thiopurines current use (e.g. azathioprine; concomitant use of any medications with ATC code L01BB, LO4AX (where Preferred name= "AZATHIOPRINE")) vs none

    - ▪ More than one of the above

- Baseline Severity Groups:

  - o Mayo score at baseline (6-8, 9-12)

  - o CRP: 6 mg/L, >6 mg/L

- Other Baseline Patient Characteristics Subgroups:

  - o Smoking status at baseline: yes or no

- Therapeutic Drug Monitoring Subgroup

  - o TDM adjustment status: yes or no

For Week 12 Clinical Remission analyses, a logistic regression model with treatment, subgroup, and the interaction of treatment-by-subgroup included as factors will be used. The treatment-by subgroup interaction will be tested at the significance level of 0.10. Treatment group differences will be evaluated within each category of the subgroup using Fisher's exact test, regardless of whether the interaction is statistically significant. Missing data will be imputed using NRI, as described in Section 6.4.1. If any group within the subgroup (for example, yes, no) is less than 10% of the total population, only descriptive statistics will be provided for that subgroup (that is, no inferential testing).

Prior Biologic UC Therapy may be summarized using ATC codes L04AA, L04AB, A07E (where preferred name="ETROLIZUMAB"). Prior Biologic use status for the purposes of subpopulation identification will be based on the 'Prior Biologic Y/N?' eCRF question.

TDM adjustment status will be determined by utilizing responses in the Exposure Based Dose Adjustment eCRF page.

Additional subgroup analyses may be performed as deemed necessary.

6.15.2. Safety Subgroup Analysis

Safety subgroup analysis for common TEAEs may be summarized by treatment and overall, for the safety population during the Induction and Maintenance Periods.

6.16. Protocol Violations

Review of all major and minor protocol violations will be performed on an ongoing basis during the conduct of the study. All protocol exemptions/violations identified from the following sources will be tracked:

- During clinical monitoring visits

- During data validation process

No patient will be excluded from the ITT population due to any protocol violations.

There is no Per-Protocol Population in this plan.

All protocol exemptions and violations that require medical review will be listed.

For the purpose of protocol deviation programming for prior medications the following imputation will be applied.

> If month and day are both missing, impute to Jan 1; and if day is missing, impute to 1.

> When year, month and day are missing the date will be treated as missing (i.e. there will be no imputations).

This imputation will be applied for partial end dates alone and will only be used for protocol deviation programmable purposes. Since many of the important protocol deviation categories identify certain medications within weeks of baseline, the end dates of prior meds are imputed as described above. Table 9 lists the categories, subcategories, and study specific terms of important protocol deviations.

Table 9 **Description of Important Protocol Deviations**

Category / Subcategory / Study_specific	Source	Methods of Identification	Comment

Category: Informed Consent			
Subcategory:			
Informed Consent Not Obtained	Non Programmable (Monitoring)	202 = Procedure done prior to obtaining ICF	
Improper Consent	Non Programmable (Monitoring)	201 = Improper administration of ICF	
Improper Consent	Non Programmable (Monitoring)	203 = Incorrect ICF version used	
Category: Eligibility			
Subcategory: Inclusion / Exclusion			
Study_specific:			
Inclusion/Exclusion	Non Programmable (Monitoring)	105 = Violation of inclusion/ exclusion criteria	
Age < 18 or > 75 years of age at the time of initial Screening	Programmable (clinical database)		Use informed consent date (this is the same as initial screening).
Have a Mayo score < 6 or > 12 (during screening*), with an endoscopic subscore < 2 in first 14 days prior to first dose of study treatment	Non Programmable (Monitoring)		Mayo is full Mayo (inclusive of PGA subscore).
Have Anti-TNF therapy within 8 weeks of baseline*	Programmable (clinical database)		As defined in section 6.15.1 of the SAP. ATC code= L04AB
Have Vedolizimab treatment within 12 weeks of baseline*	Programmable (clinical database)		Preferred_Name='VEDOLIZU MAB' and ATC code=L04AA
5-ASA compound dosing has not been stable < 2 weeks before screening*	Non Programmable (Monitoring)		ATC Code: A07EC Stable = dose level not changed.
Thiopurine prescribed dose is not stable < 8 weeks before baseline*	Programmable (clinical database)		As defined in section 6.15.1 of the SAP. Preferred name= "AZATHIOPRINE" AND ATC codes LO4AX, L01BB

Have hematologic absolute neutrophil count < 1.5 x 109/L at screening*	Programmable (clinical database)		Lbtest=' Neutrophils, Segmented'. Last lab collected on or before date of randomization will be utilized.
Have hematologic platelet count <100x109/L at screening*	Programmable (clinical database)		Lbtest=' Platelets'. Last lab collected on or before date of randomization will be utilized.
Have hematologic hemoglobin level <10.0g/dL at screening*	Programmable (clinical database)		Lbtest=' Hemoglobin' (note the unit was g/L in LB) . Last lab collected on or before date of randomization will be utilized.
Have hematologic lymphocyte count ≤500 cells/µL at screening*	Programmable (clinical database)		Lbtest=' Lymphocytes' (note the unit was 10^9/L in LB) . Last lab collected on or before date of randomization will be utilized.
Have hematologic total white blood cell count <3.0 x 109/L at screening*	Programmable (clinical database)		Leukocytes to be used for calculation of WBC count. . Last lab collected on or before date of randomization will be utilized.
Have chemistry: serum creatinine levels >2x upper limit of normal (ULN) at screening*	Programmable (clinical database)		May need to alter the ULN based on volume of subjects violating. LBTEST=' Creatinine'. Last lab collected on or before date of randomization will be utilized.
Have chemistry: total bilirubin level >2x upper limit of normal (ULN) at screening*	Programmable (clinical database)		May need to alter the ULN based on volume of subjects violating. LBTEST=' Bilirubin'. Last lab collected on or before date of randomization will be utilized.
Have chemistry: alkaline phosphatase >2x upper limit of normal (ULN) at screening*	Programmable (clinical database)		May need to alter the ULN based on volume of subjects violating. LBTEST=' Alkaline Phosphatase'. Last lab collected on or before date of randomization will be utilized.
Have chemistry: alanine aminotransferase (ALT) levels >2x upper limit of normal (ULN) at screening*	Programmable (clinical database)		May need to alter the ULN based on volume of subjects violating. LBTEST=' Alanine Aminotransferase'. Last lab collected on or before date of randomization will be utilized.

Have chemistry: aspartate aminotransferase (AST) levels >2x upper limit of normal (ULN) at screening*	Programmable (clinical database)		May need to alter the ULN based on volume of subjects violating. LBTEST=' Aspartate Aminotransferase'. Last lab collected on or before date of randomization will be utilized.
Have received cyclosporine or thalidomide ≤30 days of screening* for treatment of UC.	Programmable (clinical database)		Use Preferred name= "CICLOSPORIN" or "TACROLIMUS" when ATC Codes L04AD or Preferred name="THALIDOMIDE" when ATC code=L04AX.
Have received corticosteroid enemas, corticosteroid suppositories or topical treatment with 5-ASA ≤30 days of screening* endoscopy for treatment of UC.	Non Programmable (Monitoring)		
Have used apheresis (Adacolumn aphersis) ≤2 weeks before screening* endoscopy for treatment of UC.	Programmable (clinical database)		Use Preferred name="MEDICAL DEVICES".
Have previous exposure to any biologic therapy for IL-23 (ustekinumab), either licensed or investigational.	Programmable (clinical database)		Preferred_Name='USTEKINU MAB'.
Have abdominal abscess or toxic megacolon during screening*	Non Programmable (Monitoring)		
Have positive hepatitis B surface antigen or positive anti-hepatitis B core antibody (HBcAb+) and positive confirmatory polymerase chain reaction (PCR) for HBV, regardless of anti-hepatitis B surface antibody status at screening*	Programmable (clinical database)		'Hepatitis B Virus Surface Antigen' 'Hepatitis B Virus Core Antibody'

Have positive result for hepatitis C antibody and positive confirmatory PCR test for hepatitis C virus at screening*	Programmable (clinical database)		'Hepatitis C Virus Antibody'
Had Clostridium difficile within 30 days of screening endoscopy or tests positive at screening*	Programmable (clinical database)		

* For definition of Baseline and Screening, Date of Randomization should be used for consistency throughout.

Category: Investigational Product			
Subcategory:			
Treatment Assignment/ Randomization Error			
Dosing Error	Non Programmable (Monitoring)	104 = Study drug not administered per protocol	
Compliance	Programmable (clinical database)	Patient received a dose adjustment but was not eligible for a dose adjustment per protocol.	Only patients in the 50 mg and 200 mg induction treatment arms are eligible for adjustment.
Compliance	Programmable (clinical database)	Patient experienced a TDM dose decrease.	Protocol specified that patients will not experience a dose decrease on the basis of drug concentrations (protocol section 9.5).
Compliance (Missed Dose)	Programmable (clinical database)	Dose missed or administered outside of the protocol defined visit window.	Out of window is defined as -/+ 14 days.
Category: Study Procedures			
Subcategory:			
Excluded Conmeds	Non Programmable (Monitoring)	103 = Excluded Medication Used	
Visit Schedule Criteria	Non Programmable (Monitoring)	101 = Visit/ Procedure out of window	

Other	Non Programmable (Monitoring)	102 = Visit/ Procedure not done	
Category: Other			
Subcategory:			
Other	Non Programmable (Monitoring)	106 = Other protocol deviation	
Other	Non Programmable (Monitoring)	208 = Other regulatory deviation	
Category: Administrative/Oversight			
Subcategory:			
Improper Conduct of Assessments	Non Programmable (Monitoring)	207 = Procedure performed by non-delegated personnel	
Reg/Ethic Approvals	Non Programmable (Monitoring)	204 = IRB approval not obtained prior to screening	
Reg/Ethic Approvals	Non Programmable (Monitoring)	205 = IRB approval expired	
Category: Safety			
Subcategory:			
SAEs	Non Programmable (Monitoring)	206 = Failure/delay in SAE reporting	

6.17. Number and Timing of Analyses

6.17.1. Primary Analysis

The primary analysis will take place when all randomized patients have had an opportunity to complete 12 weeks of induction treatment.

6.17.2. Interim Analyses and Data Monitoring

One interim analysis will be conducted to assess the primary efficacy results when approximately half (~120) of the patients have completed the dosing induction phase (or discontinued study treatment). The interim efficacy results will be used for internal decision making to trigger planning activities associated with the investigational product and to aid development of PK/PD modeling. The study may not be stopped for positive efficacy. Hence, there will be no alpha adjustment the interim analysis. The assessment will be conducted by a sponsor assessment committee with a limited number of pre-identified team members who do not have direct site contact or data entry/validation responsibilities. To minimize any bias being

introduced into the analysis of the study results, the SAP and PK/PD analysis plan will be finalized and approved before the efficacy interim analysis begins.

Information that may unblind the study during the analyses will not be reported to study sites or the blinded study team until the study has been unblinded.

Ongoing monitoring of safety data (including AEs, SAEs, and selected laboratory measurements) will be continued throughout the study using blinded data. Reviewing details are specified in the trial level safety review (TLSR) plan or a separate document.

6.18. Changes to the Planned Analysis from the Clinical Study Protocol

The primary analysis method for categorical variables will now be logistic regression not CMH.

Prior Concomitant and Current Concomitant medication will not be summarized as reported in section 12.2.4 of the protocol. Prior Medication and Concomitant medications will be summarized per section 6.10 of the SAP.

As of the approval date of SAP version 3, there are no additional changes to the planned analysis in the protocol.

7. Unblinding Plan

A limited number of pre-identified individuals at Eli Lilly and Company may gain access to the unblinded data as part of the prespecified interim analysis (see section 6.17) in order to initiate the final population PK/PD model development processes for primary analyses and/or to evaluate safety data. Information that may unblind the study during the analyses will not be reported to study sites until the Maintenance period has been completed.

Additionally, a limited number of pre-identified individuals, an Eli Lilly and Company PK/PD Scientist and unblinded site pharmacists will have access to patients dosing information in order to conduct the TDM adjustments. Unblinded dosing information will not be shared with any other study or site personnel.

8. References

[1] Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.

[2] Frank-Bretz et al., Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes or parametric tests,, Biometric Journal 53 (2011) 6, 894-913. Al

[2]. Mohamed Alosh et., Advanced multiplicity adjustment methods in Clinical trials,., Statistics in Medicine, Published on line 16 September 2013 in Wiley Online Library.

[3].Mosli M et al., Development and validation of a histological index for UC, Gut 2015, 0:1-9.

[4]. Travis et al., Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS), Gut 2012, 535-42.

9.   Appendices

Appendix A. The UCEIS

The Ulcerative Colitis Endoscopic Index of Severity is evaluated at the time of endoscopy. The scoring of the sub scores can be found in Table 10.

Table 10   UCEIS Scoring Algorithm

Descriptor (score most severe lesions)	Likert scale anchor points	Definition	Total score (most severe area)
Vascular pattern	Normal (0)	Normal vascular pattern with arborisation of capillaries clearly defined, or with blurring or patchy loss of capillary margins	0
	Patchy loss (1)	Patchy obliteration of vascular pattern	
	Obliterated (2)	Complete obliteration of vascular pattern	
Bleeding	None (0)	No visible blood	1
	Mucosal (1)	Some spots or streaks of coagulated blood on the surface of the mucosa ahead of the scope, which can be washed away	2
	Luminal mild (2)	Some free liquid blood in the lumen	3
	Luminal severe (3)	Frank blood in the lumen ahead of endoscope or visible oozing from mucosa after washing intra-luminal blood, or visible oozing from a haemorrhagic mucosa	4
			5
Erosions and ulcers	None (0)	Normal mucosa, no visible erosions or ulcers	6
	Erosions (1)	Tiny (≤5mm) defects in the mucosa, of a white or yellow colour with a flat edge	7
	Superficial ulcer (2)	Larger (>5mm) defects in the mucosa, which are discrete fibrin-covered ulcers when compared to erosions, but remain superficial	8
	Deep ulcer (3)	Deeper excavated defects in the mucosa, with a slightly raised edge	

Date of preparation September 2013 AXHUG130882an

 The UCEIS score is defined as the sum of the Vascular Pattern Score (0-2), the Bleeding Score (0-3) and the Erosions and Ulcers Score (0-3). Possible values of the UCEIS score range from 0-8 with being the most severe.

Appendix B. IBDQ Missing Data Handling Algorithm

The instructions on how the IBDQ score is calculated in the presence of missing or incomplete information is provided below with detailed scenarios given in Table 11.

IBDQ: The IBDQ is a 32-item self-administered questionnaire. The IBDQ has *4 dimensions*: bowel symptoms (10 items), systemic symptoms (5 items), emotional function (12 items), and social function (5 items). Responses are graded on a *7-point Likert scale* in which 7 denotes "not a problem at all" and 1 denotes "a very severe problem." Scores range from 32 to 224; a higher score indicates a better quality of life. The 4 dimensions are defined as:

- Bowel symptoms: Questions 1, 5, 9, 13, 17, 20, 22, 24, 26, 29
- Systemic symptoms: Questions 2, 6, 10, 14, 18
- Emotional function: Questions 3, 7, 11, 15, 19, 21, 23, 25, 27, 30, 31, 32
- Social function: Questions 4, 8, 12, 16, 28

Rules for handling missing data:

1. If no response is given for a particular question and only one response per dimensional score is missing, impute the missing value to be equal to the mean score for the other items of the subscore.
2. If two or more questions are unanswered for a particular domain, then the subscore will be set to missing.
3. If after steps 1 and 2, more than 4 questions are missing for the full IBDQ, then the total IBDQ will be set to missing.

Table 11 IBDQ Missing Data Instructions

	Number of Missing values				Comments /Actions
	Subscore 1	Subscore 2	Subscore 3	Subscore 4	
Scenario A	0	0	0	0	Ok, we can compute 4 subscores and the full IBDQ
Scenario B	1	0	0	0	We can compute both the full IBDQ and the 4 subscores, by replacing the missing value with the average mean (computed on the data available for its own subscore)
Scenario C	1	1	1	1	We can compute both the full IBDQ and the 4 subscores, by replacing the missing values with the average means (each one computed on the data available for its own subscore)
Scenario D	2	0	0	0	We can compute the full IBDQ by replacing the 2 missing values with the average mean (computed on the data available for its own subscore). We can compute only 3 subscores
Scenario E	2	1	0	1	We can compute the full IBDQ by replacing the missing values with the average means (each one computed on the data available for its own subscore). We can compute only 3 subscores
Scenario F	2	1	1	1	We cannot compute the full IBDQ (more than 4 missing values). We can compute only 3 subscores
Scenario G	2	2	0	0	We cannot compute the full IBDQ because in more than one subscore there are 2 missing values (even though the total missing value is not exceeding 4). We can compute only 2 subscores.
Scenario H	0	0	3	0	We cannot compute the full IBDQ because there is one subscore with more than 2 missing values. We can compute only 3 subscores.

Appendix C. Histopathology

Components of Histological Indices

In the table 12 below the characteristics to be evaluated are listed and to which indices they belong are provided:

Table 12 Histology Components

Components	Geboes	RHI
Structural (architectural change)	X	
Chronic inflammatory infiltrate	X	X
Lamina propria eosinophils	X	
Lamina propria neutrophils	X	X
Neutrophils in epithelium	X	X
Crypt destruction	X	
Erosion or ulceration	X	X
Acute inflammatory cells infiltrate		

Geboes Score

The Geboes score assigns subjects to grades 0, 1, 2A, 2B, 3, 4 and 5 according to seven histological features: structural (architectural change), chronic inflammatory infiltrate, lamina propria eosinophils, lamina propria neutrophils, neutrophils in epithelium, crypt destruction and erosion or ulceration.

Table 13   The Geboes Grading System

Geboes Grading System
Grade 0 - Structural (architectural change)
Subgrades
0.0   No abnormality
0.1   Mild abnormality
0.2   Mild or moderate diffuse or multifocal abnormalities
0.3   Severe diffuse or multifocal abnormalities
Grade 1 - Chronic inflammatory infiltrate
Subgrades
1.0   No increase
1.1   Mild but unequivocal increase
1.2   Moderate increase
1.3   Marked increase
Grade 2A - Lamina propria eosinophils
Subgrades
2A.0   No increase
2A.1   Mild but unequivocal increase
2A.2   Moderate increase
2A.3   Marked increase
Grade 2B - Lamina propria neutrophils
Subgrades
2B.0   None
2B.1   Mild but unequivocal increase
2B.2   Moderate increase
2B.3   Marked increase
Grade 3 - Neutrophils in epithelium
Subgrades
3.0   None
3.1   < 5% crypts involved
3.2   < 50% crypts involved
3.3   > 50% crypts involved
Grade 4 - Crypt destruction
Subgrades
4.0   None
4.1   Probable - local excess of neutrophils in part of crypt
4.2   Probable - marked attenuation
4.3   Unequivocal crypt destruction
Grade 5 - Erosion or ulceration

Subgrades	
5.0	No erosion, ulceration, or granulation tissue
5.1	Recovering epithelium + adjacent inflammation
5.2	Probable erosion - focally stripped
5.3	Unequivocal erosion
5.4	Ulcer or granulation tissue

In practice, the highest grade in which there is evidence of disease is assigned. For example if <50% crypts involved (3.2) is checked and Crypt destruction is noted as 'none' (4.0) and Erosion or ulceration is 'No erosion, ulceration ,or granulation tissue' (5.0), the subject will be assigned a score or grade of 3. For scoring a subject with either abnormalities in lamina propria eosinophils or lamina propria neutrophils will be assigned a score or grade of 2.