Official Title: Phase III, Randomized, Double-Blind, Placebo-Controlled, Multicenter Study to Evaluate the Efficacy (Maintenance of Remission) and Safety of Etrolizumab Compared With Placebo in Patients With Moderate to Severe Active Ulcerative Colitis Who Are Naive to TNF Inhibitors

NCT Number: NCT02165215

Document Date: SAP Version 3: 04-May-2020

STATISTICAL ANALYSIS PLAN

TITLE: PHASE III, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTICENTER STUDY TO EVALUATE THE EFFICACY (MAINTENANCE OF REMISSION) AND SAFETY OF ETROLIZUMAB COMPARED WITH PLACEBO IN PATIENTS WITH MODERATE TO SEVERE ACTIVE ULCERATIVE COLITIS WHO ARE NAIVE TO TNF INHIBITORS

PROTOCOL NUMBER: GA29102 (Laurel)

STUDY DRUG: Etrolizumab

VERSION NUMBER: 3

IND NUMBER: 100366

EUDRACT NUMBER: 2013-004280-31

SPONSOR: F. Hoffmann-La Roche Ltd.

PLAN PREPARED BY: [redacted], M.Sc.

DATE FINAL: Version 1: 14 May 2019

DATE AMENDED: Version 2: 17 December 2019
Version 3: See electronic date stamp below.

STATISTICAL ANALYSIS PLAN AMENDMENT APPROVAL

Date and Time(UTC)	Reason for Signing	Name
04-May-2020 14:25:25	Company Signatory	[redacted]

STATISTICAL ANALYSIS PLAN AMENDMENT
RATIONALE FOR VERSION 3

- Change to Section 4.1.2.1 (Modified Intent to Treat [mITT] Population): Update the mITT Population definition to clarify it will not be used for histologic remission analysis.

- Change to Section 4.4 (Efficacy Analysis): Update to methodology for handling for treatment discontinuation intercurrent events to set to missing and remove single imputation of worst observation carried forward.

- Change to section 4.4.5 (Supplementary Analyses): Added in sensitivity analysis for key secondary endpoints calculated as applicable.

- Change to Section 4.1.1.2 (OLI Histology Evaluable Population) and Section 4.1.2.3 (Histology Evaluable Population) were made to clarify main analysis and sensitivity analysis populations.

Additional minor changes have been made to improve clarity and consistency.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDICES

GLOSSARY OF ABBREVIATIONS

AE	adverse event
CMH	Cochran–Mantel–Haenszel
CCOD	clinical cutoff date
CS	Corticosteroids
e-CRF	electronic Case Report Form
EQ-5D	EuroQoL Five-Dimension Questionnaire
Etro	Etrolizumab
ES	endoscopic subscore
HRQoL	health-related quality of life
IBDQ	Inflammatory Bowel Disease Questionnaire
ICE	intercurrent event
IS	immunosuppressants
IxRS	interactive voice/web based response system
LoPO	list of planned outputs
MCS	Mayo Clinic Score
mITT	modified intent-to-treat
mMCS	modified Mayo Clinic Score
NHI	Nancy Histology Index
OLE	open-label extension
OLI	open-label induction
PBO	Placebo
PGA	Physician's Global Assessment
PK	Pharmacokinetic
pMCS	partial Mayo Clinic Score
QOL	quality of life
RB	rectal bleeding
SAP	Statistical Analysis Plan
SF	stool frequency
SM	safety monitoring
TNF	tumor necrosis factor
UC	ulcerative colitis
UC-PRO/SS	Ulcerative Colitis Patient-Reported Outcome Signs and Symptoms

1. BACKGROUND

This Statistical Analysis Plan (SAP) describes the analyses to be performed for Study GA29102. Study GA29102 is part of a large Phase III development program for etrolizumab. Details which are common to Studies GA28950, GA29102, GA28948, GA28949, and GA29103 are described in the project SAP.

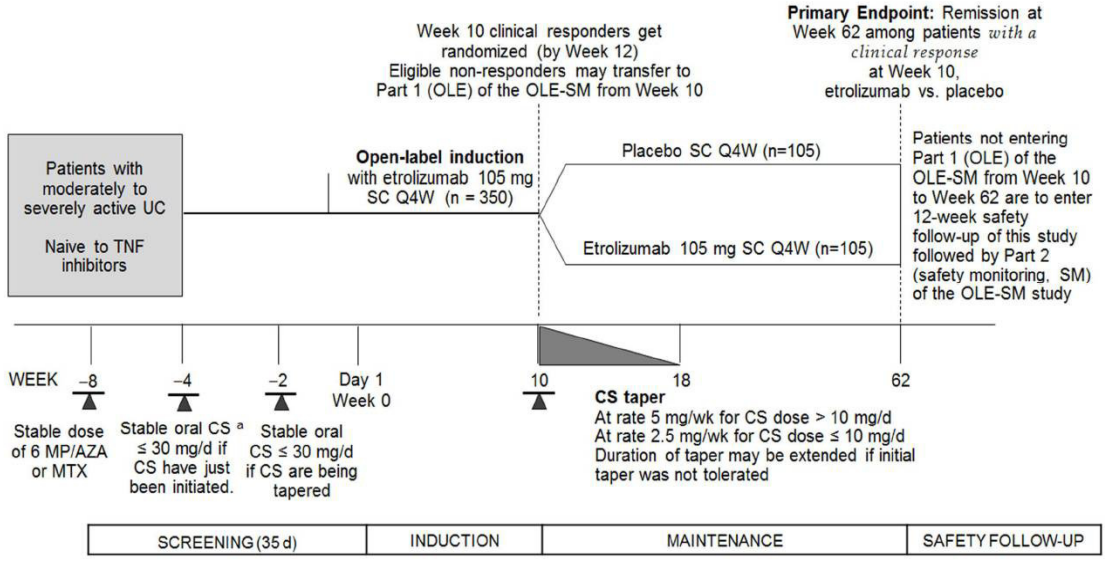
Study GA29102 is made up of two phases: the Induction Phase and the Maintenance Phase. The Induction Phase is made up of a single open-label etrolizumab arm; therefore, no statistical comparisons will be made for the Induction Phase. Induction Phase data will be summarized descriptively and pre-specified in the list of planned outputs (LoPO). Statistical analyses will only be conducted for the Maintenance Phase of the study.

The Induction and Maintenance Phases will be reported once data from the Week 62 treatment period have been collected in the database and data have been cleaned and verified; this will be referred to as the primary analysis.

2. STUDY DESIGN

This is a multicenter, Phase III, double-blind, placebo-controlled study evaluating the safety, efficacy, and tolerability of etrolizumab in patients who achieved a clinical response to etrolizumab treatment after 10 weeks. The population is made up of moderate to severe ulcerative colitis (UC) patients who are naive to tumor necrosis factor (TNF) inhibitors (see Figure 1).

Figure 1 Study Schema



6-MP=6-mercaptopurine; AZA=azathioprine; d=day; CS=corticosteroid; MTX=methotrexate; OLE=open-label extension; OLE-SM=open-label extension and safety monitoring; Q4W=every 4 weeks; QOD=every other day; SC=subcutaneous; SM=safety monitoring; TNF=tumor necrosis factor; UC=ulcerative colitis; wk=week.

[a] Stable budesonide at ≤ 9 mg/day. Taper from Week 10 to QOD for 2 weeks and then discontinue.

Following participation in this study, patients may be eligible to participate in an open-label extension and safety monitoring (OLE-SM) Study GA28951, which consists of two parts: Part 1 (designated OLE [open-label extension]) and Part 2 (designated SM [safety monitoring]).

2.1 PROTOCOL SYNOPSIS

The Protocol Synopsis is in Appendix 1. For additional details, see the Schedule of Assessments in Appendix 2.

2.2 OUTCOME MEASURES

Baseline throughout this document is defined as the last available assessment prior to first receipt of study drug in the Induction Phase. The efficacy outcome measures listed below deviate from the current Study GA29102 Protocol Version 7. The current protocol includes change from baseline in stool frequency (SF) and rectal bleeding (RB) at Week 6 as secondary outcomes measures, and endoscopic remission at Week 10, histologic disease activity change from baseline to Week 10, improvement in the endoscopic mucosa at Week 10, and change in health utilities, as assessed by the EuroQoL Five-Dimension Questionnaire (EQ-5D) from baseline to Week 10 as exploratory outcome measures. These outcomes measures cannot be analyzed due to no comparator arm at Week 6 or Week 10; therefore, they have been removed from the list below. The safety and pharmacokinetic (PK) measures can be found in the Protocol Synopsis in Appendix 1.

2.2.1 Efficacy Outcome Measures

Primary Outcome Measure

- Remission at Week 62 among patients with a clinical response at Week 10, as determined by the Mayo Clinic Score (MCS)

Secondary Outcome Measures

- Clinical remission at Week 62 among patients in clinical remission at Week 10
- Clinical remission at Week 62
- Remission at Week 62 among patients in remission at Week 10
- Improvement in endoscopic appearance of the mucosa at Week 62
- Endoscopic remission at Week 62
- Histologic remission at Week 62
- Corticosteroid-free clinical remission at Week 62 (off corticosteroids [CS] for at least 24 weeks prior to Week 62) in patients who are receiving corticosteroids at baseline
- Corticosteroid-free remission at Week 62 (off corticosteroids for at least 24 weeks prior to Week 62) in patients who are receiving corticosteroids at baseline
- Change from baseline to Week 62 in UC bowel movement signs and symptoms as assessed by the Ulcerative Colitis Patient-Reported Outcome Signs and Symptoms (UC-PRO/SS) measure

- Change from baseline to Week 62 in UC functional symptoms as assessed by the UC-PRO/SS measure
- Change from baseline to Week 62 in patients' health-related quality of life (HRQoL) as assessed by the overall score of the Inflammatory Bowel Disease Questionnaire (IBDQ)

Exploratory Efficacy Outcome Measures

- Clinical response at Week 62 among patients with a clinical response at Week 10
- Remission at Week 62 among patients in clinical remission at Week 10
- Corticosteroid-free clinical remission at Week 62 (off corticosteroids for at least 12 weeks prior to Week 62) in patients who are receiving corticosteroids at baseline
- Histologic disease activity change from baseline to Week 62
- Improvement in histologic and/or endoscopic disease activity
- Change in health utilities, as assessed by the EQ-5D, from baseline to Week 62
- Frequency and duration of hospitalizations from Week 10 to Week 62
- Response, remission and corticosteroid-free endpoints as determined by the modified Mayo Clinic Score (mMCS)

2.3 DETERMINATION OF SAMPLE SIZE

2.3.1 Sample Size

The study sample size was selected so that sufficient patients are enrolled to evaluate the primary endpoints in the Maintenance Phase. Approximately 350 patients will be enrolled in the open-label induction (OLI) arm and, under the assumption of 60% of patients being clinical responders at Week 10, approximately 210 patients (previously treated with etrolizumab in the OLI phase) will be randomized into the Maintenance Phase.

The primary endpoint is Week 62 remission among patients in clinical response at Week 10. A sample size of 105 patients per arm in the Maintenance Phase will provide $>$ 90% power to detect a 25% absolute difference in remission rates between the etrolizumab and placebo Maintenance arms for a chi-squared test at the 5% significance level, under the assumption of a placebo true remission rate 30%–45%.

2.4 ANALYSIS TIMING

The primary analysis for the Induction and Maintenance Phases will be conducted once the last patient has completed their Week 62 treatment period or has discontinued from study treatment. The date associated with this visit will be termed the clinical cutoff date (CCOD) for the primary analysis, and data in the study database to the CCOD will be cleaned and verified. Sponsor personnel will be unblinded to treatment assignment to perform the primary analysis. Patients and study site personnel will remain blinded to individual Maintenance Phase treatment assignment until after the study is completed

(after all patients have either completed the safety follow-up periods or discontinued early from the study) and the database is locked.

Additional analyses will be conducted once all patients have completed 12 weeks safety follow-up and the database has been locked. These analyses will report all adverse events (AEs) including the data collected in the 12-week safety follow–up period.

3. STUDY CONDUCT

3.1 RANDOMIZATION

Details of the randomization process are included in the project SAP, Section 3.1. The statistical analyses will be conducted using the stratification factors entered into the interactive voice/web based response system (IxRS) system at randomization. Sensitivity analysis of the primary endpoints using data collected in the clinical database will be conducted as required, if the database does not match IxRS due to incorrect stratification data being collected in IxRS at randomization. Due to low or zero counts in any one stratum, the stratification factors may need to be combined or removed in the primary analysis to ensure the Cochran–Mantel–Haenszel (CMH) test is not invalidated. This approach has been applied to the Maintenance Phase as described in Section 3.1.1, however further combining or removal maybe re required to ensure the analysis is not invalidated due to zero counts.

3.1.1 Maintenance Phase

Patients who achieved a clinical response at Week 10 and did not receive (permitted/non-permitted) rescue therapy during the Induction Phase were eligible to be randomized into the Maintenance Phase following enrolment into the Induction Phase. Patients were randomized to etrolizumab or placebo using a 1:1 ratio and a permuted blocks stratified randomization (dynamically generated). Patients were stratified using the following stratification factors:

- Concomitant treatment with CS at baseline (yes/no)
- Concomitant treatment with immunosuppressants (IS) after Week 10 (yes/no)
- Disease activity at screening MCS ≤9/MCS≥10)
- Remission criteria at week 10 (No remission/ remission including Physician's Global Assessment (PGA) subscore /remission not including PGA subscore)

The Sponsor conducted a blinded review of the study data and concluded zero or low counts are likely to occur within multiple stratum when using the stratification factor 'Remission Criteria at Week 10'. This scenario would invalidate the CMH test, the Sponsor proposes not to fit 'Remission Criteria at Week 10' as a term within the analysis models.

All maintenance analysis will, therefore, be adjusted for using the following three stratification factors:

- Disease activity at screening (MCS ≤9/MCS≥10)
- Concomitant treatment with CS at baseline (yes/no)
- Concomitant treatment with IS after Week 10 (yes/no)

3.2 INDEPENDENT REVIEW FACILITY

Details are included in the project SAP, Section 3.2.

3.3 DATA MONITORING

Details are included in the project SAP, Section 3.3.

4. STATISTICAL METHODS

4.1 ANALYSIS POPULATIONS

4.1.1 Induction Phase

4.1.1.1 Open-Label Induction (OLI) Population

Patients enrolled into the Induction Phase who received study drug will be summarized under the OLI population for all efficacy summaries.

4.1.1.2 OLI Histology Evaluable Population

The histology evaluable patients will include all patients in the OLI population who provided a baseline histology sample. Patients with absence of baseline neutrophilic inflammation defined by a Nancy Histology Index (NHI) ≤ 1 will be excluded.

4.1.1.3 Safety Population

The safety analysis population for the Induction Phase will include all patients who received at least one dose of etrolizumab during the Induction Phase.

4.1.2 Maintenance Phase

4.1.2.1 Modified Intent-to-Treat (mITT) Population

Efficacy analyses for the Maintenance Phase will be performed using a modified intent-to-treat (mITT) analysis population (with the exception of histologic remission). This analysis population will include patients randomized in the Maintenance Phase who received at least one dose of maintenance study drug. Patients will be grouped according to the treatment assigned at maintenance randomization.

To allow consistency across the program the mITT population is being used. This differs to terminology in the Protocol Version 7, Section 6.4, which uses intent-to-treat population.

4.1.2.2 Pharmacokinetic (PK)-Evaluable Population

The PK-evaluable population includes patients who have received at least one dose of study drug and have at least one quantifiable concentration measured during the Maintenance Phase.

4.1.2.3 Histology-Evaluable Population

The histology-evaluable patients will include all patients in the maintenance mITT population, who have documented neutrophilic inflammation at baseline. For the main analysis based on NHI, neutrophilic inflammation is characterized by NHI greater than 1. Sensitivity analyses will be based on the Robarts Histopathological Index (RHI) and Geboes Grading Scale. Definitions for neutrophilic inflammation and histologic remission under these alternative histologic scoring systems are provided in Table 3 of the project SAP.

4.1.2.4 Safety Population

The safety analysis population for the Maintenance Phase will include all patients who receive study drug in the Maintenance Phase. Patients will be included in the treatment arm for the treatment most frequently received during the Maintenance Phase.

4.2 ANALYSIS OF STUDY CONDUCT

All data available in the database up to the point the last patient has completed their Week 62 visit will be included to evaluate study conduct. This will include all available data from the Induction and Maintenance Phases at the time of cutoff for the primary analysis. The following analyses will be performed to evaluate the study conduct:

- Summary of protocol deviations
- Summaries of mITT, histology-evaluable, PK-evaluable, and safety populations, including numbers of patients in each population
- Summary of patient disposition, including the number of doses received, reasons for patients withdrawing from the study and from study treatment, and number of patients taking study treatment.

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Details are included in the project SAP, Section 4.3.

4.4 EFFICACY ANALYSIS

All statistical hypotheses for the primary and key secondary endpoints will be evaluated under a multiple testing procedure to ensure an overall type I error no greater than 5%. Details on this testing procedure are provided in Section 4.4.3.

Statistical comparisons will only be conducted in the Maintenance Phase. Efficacy data collected during the Induction Phase will be summarized descriptively.

All comparisons between the etrolizumab and placebo arms for binary data will use the CMH test statistic stratified by the three IxRS stratification factors (Disease activity at screening [MCS ≤9/MCS≥10], Concomitant treatment with CS at baseline [yes/no], Concomitant treatment with IS after Week 10 [yes/no]) as described in Section 3.1.1. For all analyses in the Induction and Maintenance Phases, where comparisons back to baseline are made, baseline is defined as the last available assessment prior to first receipt of study drug in the Induction Phase.

In alignment with the addendum to ICH E9, the primary efficacy treatment effect (estimands) is described for the primary treatment effect in Section 4.4.1.1. The estimand attributes include the following two intercurrent events: treatment discontinuation and use of rescue medication as described below:

- **Treatment discontinuation:** All patients who discontinue study drug during the treatment period of the Maintenance Phases will not be assessed for any future Week 62 endpoints. Therefore, these patients will be assumed to be non-responders in the categorical endpoints analyses. For continuous endpoints (e.g., IBDQ, UC-PRO/SS) these patients' data will be set to missing. Handling of missing data is described in the project SAP, Section 4.8.
- **Use of rescue therapy:** Increased or new background medications compared to baseline for the treatment of UC is considered rescue therapy. Rescue therapy use is described in the protocol and also summarized in Appendix 3. All patients receiving permitted rescue therapy during the Maintenance Phases will be asked to continue the study through endpoint assessment and safety follow-up. All patients receiving prohibited rescue therapy during the Maintenance Phases will be asked to enter safety follow-up, with no assessment of any future Week 62 endpoints. Patients who receive rescue therapy (be it permitted or prohibited) during the Induction Phase are not eligible to enroll in the Maintenance Phase. Patients who receive (permitted or prohibited) rescue medication will be considered non-responders in the primary analysis for all time points following the time they received rescue therapy. For continuous outcomes (e.g., IBDQ, UC-PRO/SS), scores collected after the first use of rescue medication will have their data imputed using the worst post-baseline score from the following assessments: (the last score available prior to the start date of first rescue medication and all scores available after the start date of rescue medication use).

4.4.1 Primary Efficacy

4.4.1.1 Primary Treatment Effect

The primary efficacy treatment effect (estimand) targeted is described by the following four attributes:

a) **Population:** Adult patients with moderate to severe active UC who are naive to TNF inhibitors

b) **Variable:** Remission at Week 62 given the patient is an etrolizumab-induction phase clinical responder at Week 10.

c) **Intercurrent Event (ICE):**

- Treatment discontinuation
- Use of rescue therapy

Details of the ICE strategy are explained in Section 4.4.

d) **Population-Level Summary:** Difference in proportion of patients between etrolizumab and placebo treatment groups in the Maintenance Phase.

4.4.1.2 Primary Efficacy Endpoint

The primary efficacy endpoint is the difference in the proportion of patients in remission at Week 62 between the etrolizumab and placebo patients who achieved clinical response at Week 10.

$$\textbf{Proportion of patients in remission at Week 62 among clinical responders at Week 10} = \frac{\#\textbf{of patients in remission at Week 62}}{\#\ \textbf{of patients in Maintenance mITT}}$$

The calculation will be done within treatment groups. Comparisons will only be made between the etrolizumab patients and placebo patients included in the mITT population.

Null hypothesis H0: ρ Etro – ρ PBO = 0; the proportion of patients achieving remission at Week 62 among clinical responders at Week 10 in the placebo arm is the same as the proportion of patients achieving remission among clinical responders in the respective etrolizumab arm.

Alternative hypothesis H1: ρ Etro – ρ PBO ≠ 0; the proportion of patients achieving remission at Week 62 among clinical responders at Week 10 in the placebo arm is not the same as the proportion of patients achieving remission among clinical responders in the respective etrolizumab arm.

4.4.1.3 Remission Among Clinical Responders Definition

Details of remission and clinical responder definition are provided in the project SAP, Section 4.4.2.1.4.

4.4.2 Secondary Efficacy Endpoints

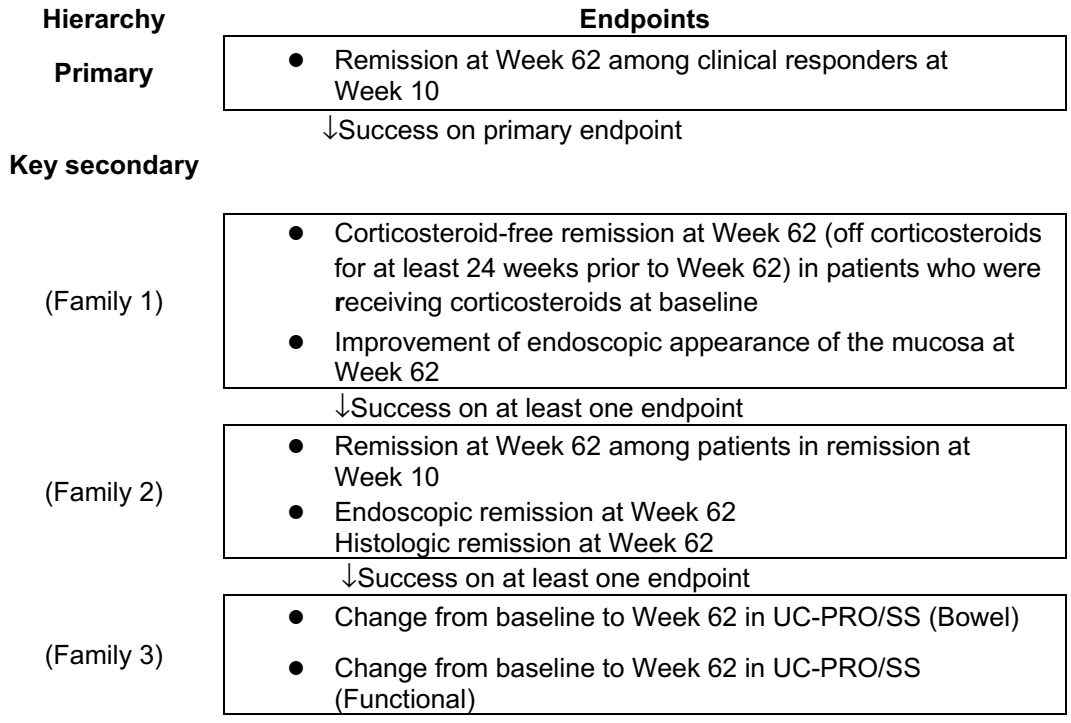
4.4.2.1 Control of Type I Error

The primary and key secondary endpoints (Figure 2) will be evaluated in a hierarchical manner with multiplicity control via multistage gatekeeping under the truncated Holm multiple testing procedure (Dmitrienko et al. 2008). In Families 1 and 2 of the key secondary endpoints the same truncation value of 0.5 will be applied; no truncation is applied in Family 3, since it is prioritized lowest in the hierarchy and imposes no gatekeeping restrictions on subsequent families. The truncation parameter value and the relative effect sizes of the endpoints influence how power is balanced over the secondary endpoint families. Truncation parameter values close to 1 should generally

be chosen when the effect sizes in high-priority endpoints are relatively large in order to maximize overall power. In other settings where the effect sizes might be smaller or mixed, a truncation parameter value closer to 0 could improve overall power. A parameter value strictly less than one in testing families that impose parallel gatekeeping (i.e. at least one endpoint is considered statistically significant after multiplicity adjustment in order to test the next family) is needed to achieve strong overall type I error control (Dmitrienko et al. 2009).

Under this multistage gatekeeping approach, multiplicity-adjusted p-values will reflect the gatekeeping restrictions depicted in Figure 2. In particular, the adjusted p-values for all key secondary endpoints in Families 1, 2, and 3 will be no smaller than the p-value for the primary endpoint. Similarly, the smallest adjusted p-value in Family 2 (3) will be no smaller than the smallest adjusted p-value in Family 1 (2). Endpoints for which the multiplicity-adjusted p-value is greater than 5% will not be considered statistically significant. These endpoints and all endpoints not under multiplicity control will be considered to provide supportive information. By repeated application of Proposition 4.1 of Dmitrienko et al. (2008), this testing strategy ensures that the overall type I error is no greater than 5%.

Figure 2 Multiple Testing Procedure for Endpoints



RB = rectal bleeding; SF = stool frequency; UC-PRO/SS = Ulcerative Colitis Patient-Reported Outcome Signs and Symptoms.

Ordering of endpoints within a family will be based on the p-value results from the hypotheses tests of the endpoints in Figure 2.

4.4.2.2 Corticosteroid-Free Remission at Week 62 (Off Corticosteroid for At Least 24 Weeks Prior to Week 62) in Patients Who Were Receiving Corticosteroids at Baseline

This analysis will only be conducted on a subgroup of patients who were receiving corticosteroids at baseline as per the data collected in electronic Case Report Form (e-CRF) database and were enrolled into the Maintenance Phase. Remission is determined by the MCS defined in the project SAP, Section 4.4.2.1.4. Patients are defined as being off CS if they have no record of taking CS since the date which is 24 weeks prior to Week 62. This date is defined as the Week 62 visit date−168 days.

$$\begin{aligned}&\textbf{Proportion of patients in corticosteroids free remission at Week 62}\\&\textbf{(off corticosteroid for at least 24 weeks prior to Week 62)}\\&\textbf{in patients who were receiving corticosteroids at baseline}\\&=\frac{\textbf{\# of patients in remission at Week 62 \& off CS for 24 weeks prior to Week 62}}{\textbf{\# of patients receiving CS at Baseline in the maintenence mITT}}\end{aligned}$$

4.4.2.3 Improvement in Endoscopic Appearance of the Mucosa

The proportion of patients with improvement in endoscopic appearance of the mucosa at Week 62 will be analyzed using the same methods as the primary endpoint. Definition of improvement in the appearance of the endoscopic mucosa is provided in the project SAP, Section 4.4.2.

4.4.2.4 Remission at Week 62 (Among Patients in Remission at Week 10)

The proportions of patients who achieved remission at Week 62 given they achieved remission at Week 10 will be analyzed using the same methods as the primary endpoint, and include patients randomized into the maintenance mITT population. Due to low number of patients in this analysis the stratification factor IS use after Week 10 will not be included in the model, to prevent any strata with no values impacting the analysis. The definition of remission is the same at each time point and is described in project SAP, Section 4.4.2.1.4.

$$\begin{aligned}&\textbf{Proportion of patients in remission at Week 62 among patients in remission at Week 10}\\&=\frac{\textbf{\# of patients in remission at Week 10 and Week 62}}{\textbf{\# of patients in remission at Week 10}}\end{aligned}$$

4.4.2.5 Endoscopic Remission

The proportion of patients with endoscopic remission at Week 62 will be analyzed using the same methods as the primary endpoint. Definition of endoscopic remission is provided in the project SAP, Section 4.4.2.1.4.

4.4.2.6 Histologic Remission

The proportion of patients with histologic remission at Week 62 will be analyzed using the same methods as the primary endpoint using the histology evaluable population.

Definition of histologic remission is described in the project SAP, Section 4.4.2.2. Further details of additional analysis to support this endpoint using different scoring systems are detailed in the project SAP, Section 4.4.2.2.

4.4.2.7 UC-PRO/SS

Details of UC-PRO/SS endpoints and the analyses are provided in the project SAP, Section 4.4.2.3. The change from baseline at Week 62 in both the functional and bowel domains will be analyzed separately.

4.4.2.8 Corticosteroid-Free Clinical Remission at Week 62 (Off Corticosteroid for At Least 24 Weeks Prior to Week 62) in Patients Who Were Receiving Corticosteroids at Baseline

This endpoint will be analyzed using the same methods as Section 4.4.2.2. Remission will be substituted with Clinical Remission. Clinical Remission is determined by the MCS and described in the project SAP, Section 4.4.2.1.4. As detailed in Figure 2 (Section 4.4.2.1), this endpoint is not included in the multiple testing procedure for endpoints.

4.4.2.9 Clinical Remission at Week 62 (Among Patients in Clinical Remission at Week 10)

The endpoint will be analyzed using the same methods as Section 4.4.2.4. Remission will be substituted with clinical remission. Clinical Remission is determined by the MCS and described in the project SAP, Section 4.4.2.1.4. As detailed in Figure 2 (Section 4.4.2.1), this endpoint is not included in the multiple testing procedure for endpoints.

4.4.2.10 Clinical Remission

Definition of clinical remission is provided in the project SAP, Section 4.4.2.1.4. The proportion of patients in clinical remission at Week 62 will be analyzed using the same methods as the primary endpoint. As detailed in Figure 2 (Section 4.4.2.1), this endpoint is not included in the multiple testing procedure for endpoints.

4.4.2.11 Clinical Response

Definition of clinical response is provided in the project SAP, Section 4.4.2.1.4. The proportion of patients in clinical response at Week 62 will be analyzed using the same methods as the primary endpoint. As detailed in Figure 2 (Section 4.4.2.1), this endpoint is not included in the multiple testing procedure for endpoints.

4.4.2.12 Inflammatory Bowel Disease Questionnaire (IBDQ)

Details of the IBDQ endpoint and the analyses are provided in the project SAP, Section 4.4.2.4. The change from baseline at Week 62 in IBDQ will be analyzed. As detailed in Figure 2 (Section 4.4.2.1), this endpoint is not included in the multiple testing procedure for endpoints.

4.4.3 Exploratory Efficacy Endpoints

Exploratory endpoints detailed in the protocol will also be analyzed. Binary endpoints will use the same statistical methods used for the primary endpoint. Continuous endpoints will use the same statistical methods used for the change from baseline key secondary endpoints.

4.4.4 Sensitivity Analyses

To support the primary analyses the following sensitivity analyses will be conducted: definitions of partial Mayo Clinic Score (pMCS) and mMCS are included in the project SAP, Section 4.4.2.1.4.

- Proportion of patients in remission at Week 62 among clinical responders at Week 10 (remission derived using mMCS including mild friability in Mayo ES = 1)
- Proportion of patients in remission at Week 62 among clinical responders at Week 10 (remission derived using pMCS)
- Proportion of patients in remission at Week 62 among clinical responders at Week 10 (difference in proportions calculated using Fisher's exact test)
- Proportion of patients in remission (mMCS) at Week 62 among clinical responders (mMCS) at Week 10 (excluding friability from Mayo ES = 1)
- Proportion of patients in remission at Week 62 among clinical responders at Week 10 (Tipping Point Analysis)

All key secondary endpoints included in Figure 2 which are calculated using MCS or the endoscopic subscore will be re analyzed as a sensitivity analyses using the modified Mayo Clinic Score (mMCS) excluding friability from ES = 1 or the endoscopic subscore excluding friability from ES = 1 if applicable. This analysis will be conducted on the mITT population and are listed below.

- Proportion of patients corticosteroid free remission at Week 62 (Off corticosteroid for at least 24 weeks prior to Week 62) in patients who were receiving corticosteroids at baseline (remission derived using the mMCS excluding mild friability in Mayo ES = 1)
- Proportion of patients with improvement of endoscopic appearance of the mucosa at Week 62 (endoscopic score excluding mild friability in Mayo ES = 1)
- Proportion of patients remission at Week 62 among patients in remission at Week 10 (remission at Week 10 and Week 62 derived using the mMCS excluding mild friability in Mayo ES = 1)

4.4.5 Supplementary Analyses

To support the primary analyses the following supplementary analyses will be conducted;

- Proportion of patients in remission at Week 62 among clinical responders at Week 10 (including data collected whilst patients received rescue therapy)

- Proportion of patients in remission at Week 62 among clinical responders at Week 10 (using logistic regression model)
- Proportion of patients in remission at Week 62 among clinical responders at Week 10 (using clinical database derivation of clinical response at Week 10, if appropriate)

Definitions of clinical response are included in the project SAP, Section 4.4.2.1.4.

4.4.6 Subgroup Analyses

Subgroup analyses are detailed in the project SAP, Section 4.4.3.

4.5 PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Details are included in the project SAP, Section 4.5.

4.6 SAFETY ANALYSES

Data reported during the Induction Phase using the induction safety population will include data for all patients enrolled into the Induction Phase, up until they are randomized into the Maintenance Phase. For patients who are not randomized into the Maintenance Phase, data will be reported up until the patient completes/withdraws from the study. If a patient is ongoing in the Induction Phase due to being in safety follow-up at the time of the primary data analyses, all available data at time of the primary analyses CCOD will be reported.

Data reported during the Maintenance Phase will include data for all patients in the maintenance safety population up until the patient complete/withdraws from the study. All safety analyses using the maintenance safety population data will be reported from baseline. Baseline is defined as the last available assessment prior to first receipt of study drug at the beginning of the Induction Phase. If a patient is ongoing in the Maintenance Phase due to being in safety follow-up at the time of the primary data analyses, all available data at time of the primary analyses CCOD will be reported.

Further details of the safety analyses are included in the project SAP, Section 4.7.

4.6.1 Exposure of Study Medication

The number of doses of etrolizumab or placebo injected will be summarized using descriptive statistics.

4.6.2 Adverse Events

Details are included in the project SAP, Section 4.7.1.

4.6.3 Laboratory Data

Details are included in the project SAP, Section 4.7.2.

4.6.4 Vital Signs

Details are included in the project SAP, Section 4.7.3.

4.6.5 Concomitant Medications

Details are included in the project SAP, Section 4.7.5.

4.6.6 Medical History

Details are included in the project SAP, Section 4.7.4.

4.7 MISSING DATA

Details are included in the project SAP, Section 4.8.

4.8 INTERIM ANALYSES

No efficacy interim analyses are planned or have been undertaken.

5. REFERENCES

Dmitrienko A, Tamhane AC, Wiens BL. General multistage gatekeeping procedures. Biom J. 2008;50(5):667–77.

Dmitrienko A, Tamhane AC, Bretz F, editors. Multiple testing problems in pharmaceutical statistics. CRC Press 2009;165–91.

Appendix 1
Protocol Synopsis

TITLE:	**PHASE III, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTICENTER STUDY TO EVALUATE THE EFFICACY (MAINTENANCSE OF REMISSION) AND SAFETY OF ETROLIZUMAB COMPARED WITH PLACEBO IN PATIENTS WITH MODERATE TO SEVERE ACTIVE ULCERATIVE COLITIS WHO ARE NAIVE TO TNF INHIBITORS**
PROTOCOL NUMBER:	GA29102
VERSION NUMBER:	7
EUDRACT NUMBER:	2013-004280-31
IND NUMBER:	100366
TEST PRODUCT:	Etrolizumab (PRO145223, RO5490261)
PHASE:	III
INDICATION:	Ulcerative colitis
SPONSOR:	F. Hoffmann-La Roche Ltd

Objectives

Efficacy Objectives

The primary efficacy objective for this study is as follows:

- To evaluate the efficacy of etrolizumab (105 mg subcutaneous [SC] every 4 weeks [Q4W]) compared with placebo remission at Week 62 among patients with a clinical response at Week 10, as determined by the Mayo Clinic Score (MCS)

The secondary efficacy objectives for this study are as follows:

- To evaluate maintenance of clinical remission at Week 62 for patients in clinical remission at Week 10
- To evaluate clinical remission at Week 62
- *To evaluate remission at Week 62 among patients in remission at Week 10*
- To evaluate improvement in endoscopic appearance of the mucosa at Week 62
- To evaluate endoscopic remission at Week 62
- To evaluate histologic remission at Week 62
- *To evaluate onset of action, defined as change from baseline in rectal bleed subscore at Week 6*
- *To evaluate onset of action, defined as change from baseline in stool frequency subscore at Week 6*
- To evaluate corticosteroid-free clinical remission at Week 62 (off corticosteroids for at least 24 weeks prior to Week 62) in patients who were receiving corticosteroids at baseline
- To evaluate corticosteroid-free remission at Week 62 (off corticosteroids for at least 24 weeks prior to Week 62) in patients who were receiving corticosteroids at baseline
- To evaluate change from baseline to Week 62 in UC bowel movement signs and symptoms, as assessed by the Ulcerative Colitis Patient-Reported Outcome Signs and Symptoms (UC-PRO/SS) measure

- To evaluate change from baseline to Week 62 in UC abdominal symptoms, as assessed by the UC-PRO/SS measure
- To evaluate change from baseline to Week 62 in patient-reported health-related QOL, as assessed by the Inflammatory Bowel Disease Questionnaire (IBDQ)

The exploratory efficacy objectives for this study are as follows:

- *To evaluate clinical response at Week 62 among patients with a clinical response at Week 10*
- *To evaluate remission achieved at Week 62 among patients in clinical remission at Week 10*
- To evaluate corticosteroid-free clinical remission at Week 62 (off corticosteroids for at least 12 weeks prior to Week 62) in patients who were receiving corticosteroids at baseline
- *To evaluate change in histologic disease activity from baseline to Week 10 and Week 62*
- *To evaluate improvement in histologic and/or endoscopic disease activity*
- To evaluate change in health utilities, as assessed by the EuroQoL Five-Dimension Questionnaire (EQ-5D), from Week 10 to Week 62
- To evaluate improvement in endoscopic appearance of mucosa at Week 10
- To evaluate endoscopic remission at Week 10
- To evaluate the frequency and duration of hospitalizations from Week 10 to Week 62
- *To evaluate response, remission, and corticosteroid-free endpoints, as determined by the mMCS*

Safety Objectives

The safety objectives for this study are as follows:

- To evaluate the overall safety and tolerability of etrolizumab over a period of 62 weeks
- To evaluate the incidence and severity of infection-related adverse events
- To evaluate the incidence of malignancies
- To evaluate the incidence and severity of hypersensitivity reactions
- To evaluate the incidence and the clinical significance of anti-therapeutic antibodies (ATAs)

Pharmacokinetic Objectives

The pharmacokinetic (PK) assessment will be performed in all patients during the Induction Phase and in all patients who were randomized into the Maintenance Phase.

The PK objectives for this study are as follows:

- To evaluate etrolizumab serum concentration at the time of primary endpoint evaluation (Week 62) and predose concentration at the steady state during the maintenance dosing period
- To evaluate the interindividual variability and potential covariate effects on etrolizumab serum exposure

Exploratory Pharmacodynamic and Diagnostic Objectives

The exploratory pharmacodynamics (PD) and diagnostic objectives for this study are as follows:

- To evaluate the relationship between baseline colonic mucosal biomarkers and/or peripheral blood and response to study drug, including, but not limited to, the αE integrin
- To evaluate the levels of biomarkers in colonic tissue and/or peripheral blood at baseline and during the treatment period, including, but not limited to, the αE integrin
- To evaluate the PD effects on biomarkers in colonic tissue and/or peripheral blood following study treatment or placebo
- To evaluate biomarkers in stool at baseline and during the treatment period through assessments that may include, but are not limited to, analyses of the microbiota and bacterial cultures

Study Design

Description of Study

This is a multicenter, Phase III, randomized, double-blind, parallel-group study to evaluate the safety, efficacy, and tolerability of etrolizumab (105 mg SC Q4W) compared with placebo in the treatment of UC. Safety and efficacy of continued etrolizumab treatment will be evaluated in patients achieving a clinical response after 10 weeks.

Patients enrolled in this study may be eligible to participate in an open-label extension and safety monitoring (OLE-SM) study (GA28951), which consists of two parts: Part 1 (designated OLE [open-label extension]) and Part 2 (designated SM [safety monitoring]).

Disease severity will be measured using the MCS, which is the current outcome measure accepted by regulatory authorities for drug development in UC. The target population is patients with moderately to severely active UC (defined as MCS 6−12, endoscopy subscore of $\geq$2 as determined by the central reading procedure described in the protocol, a rectal bleeding subscore $\geq$1, and a stool frequency subscore $\geq$1) and involvement that extends a minimum of 20 cm from the anal verge.

All patients are to be naive to tumor necrosis factor (TNF) inhibitors.

Patients must have had an inadequate response, loss of response, or intolerance to prior immunosuppressant and/or corticosteroid treatment.

Patients who are on background immunosuppressant therapy (6-MP, AZA, MTX) may be enrolled if they have received a stable dose for at least 8 weeks prior to Day 1. Such patients should continue on stable doses of their background immunosuppressant therapy during the study unless dose reduction or discontinuation is required due to toxicity.

Generally accepted criteria for discontinuation of immunosuppressants due to toxicity include, but are not limited to, acute pancreatitis, severe leukopenia, severe thrombocytopenia, or clinically significant elevations of the liver-associated enzymes from baseline, especially in the presence of an elevated total bilirubin. The ultimate decision to reduce dose or discontinue immunosuppressants due to toxicity remains at the discretion of the investigator.

Patients on oral corticosteroid therapy (prednisone at a stable dose of $\leq$30 mg, or equivalent) may be enrolled according to the following criteria:

- If corticosteroid therapy is ongoing or has recently been initiated, the dose has to be stable for at least 4 weeks immediately prior to Day 1
- If corticosteroids are being tapered, the dose has to be stable for at least 2 weeks immediately prior to Day 1

Such patients should continue stable doses of their background corticosteroid until Week 10, at which point a corticosteroid taper will be initiated.

Initiation of corticosteroid or an increase in corticosteroid dose above the patients' entry dose (up to a maximum of 30 mg/day prednisone [or equivalent]) will not be permitted during screening. Use of budesonide will be allowed at stable doses ($\leq$9 mg) provided that the dose has been stable for $\geq$4 weeks prior to Day 1. Oral 5-ASA treatment and probiotics for the treatment of UC may be continued at a stable dose as long as the dose(s) had been stable for $\geq$4 weeks and $\geq$2 weeks, respectively, prior to Day 1. Certain concomitant treatments are prohibited (see the protocol for list of all prohibited concomitant treatments).

A total of approximately 350 patients will be recruited (see the protocol) from approximately 125 sites. The study will be divided into:

- Screening period of up to 35 days
- Open-label etrolizumab treatment period of 10 weeks (Induction Phase)
- Double-blind treatment period of 52 weeks (Maintenance Phase)
- Safety follow-up period of 12 weeks.

All patients entering the study will receive induction treatment with open-label etrolizumab (105 mg SC Q4W) for 10 weeks. Eligibility for entry into the Maintenance Phase will be determined between Weeks 10 and 12. Patients who achieved a clinical response at Week 10 (see the protocol for definition of clinical response) will be randomized in a 1:1 ratio by Week 12 into the Maintenance Phase and will receive either etrolizumab (105 mg SC Q4W) or placebo

(n≈105 for each arm). Randomization will be stratified by their remission status at Week 10, concomitant treatment with corticosteroids (including budesonide) at baseline, concomitant treatment with immunosuppressants after Week 10, and disease activity measured during screening (MCS ≤ 9/MCS ≥ 10).

Patients not achieving clinical response at Week 10 and patients completing the Maintenance Phase at Week 62 should enroll in Part 1 (OLE) of Study GA28951, if eligible, where they will receive open-label etrolizumab. If they do not enroll in Part 1 (OLE) of Study GA28951, they will enter the 12-week safety follow-up period of this study and then be requested to enroll in Part 2 (SM) of Study GA28951 for 92 weeks of extended for progressive multifocal leukoencephalopathy (PML) monitoring.

Number of Patients

Approximately 350 patients are to be enrolled for this study. Following 10 weeks of open-label induction with etrolizumab, it is estimated that approximately 60% of patients will achieve clinical response, which will provide approximately 105 patients in each treatment arm for the Maintenance Phase. *Additional patients may be enrolled in the Induction Phase if needed to meet sample size requirements for the Maintenance Phase.*

Target Population

Patients must meet the following criteria for study entry:

- Able and willing to provide written informed consent
- 18–80 years of age, inclusive
- Diagnosis of UC established at least 3 months prior to Day 1 by clinical and endoscopic evidence. This diagnosis should be corroborated by histopathology conducted at any time prior to screening and documented by a histopathology report. (Note that histopathology may be performed at screening if no prior report is readily available.)
- Moderately to severely active UC as determined by an MCS of 6–12 with an endoscopic subscore ≥ 2 as determined by the central reading procedure, a rectal bleeding subscore ≥ 1, and a stool frequency subscore ≥ 1 during the screening period (prior to Day 1)
- Evidence of UC extending a minimum of 20 cm from the anal verge as determined by baseline endoscopy (flexible sigmoidoscopy or colonoscopy) performed during screening, 4–16 days prior to Day 1. See protocol for additional information regarding the time window.
- Naive to treatment with TNF inhibitors
- Patients must have had an inadequate response, loss of response, or intolerance to prior immunosuppressant and/or corticosteroid treatment.

Inadequate response to, loss of response to, or intolerance to prior immunosuppressant treatment is defined as one or more of the following:

Persistent signs and symptoms of active disease despite a history of at least one 12-week regimen of oral AZA (≥ 1.5 mg/kg) or 6-MP (≥ 0.75 mg/kg) and/or MTX (≥ 15 mg/week) within the previous 5 years

Persistent signs and symptoms of active disease despite a 6-TG level of ε230 pmol/8 $\cdot 10^8$ RBCs during at least one 12-week regimen of oral AZA or 6-MP at a stable or increasing dose within the previous 5 years

History of intolerance to AZA, 6-MP, or MTX (including, but not limited to, nausea/vomiting, abdominal pain, pancreatitis, LFT abnormalities, lymphopenia, TPMT genetic mutation, infection) within the previous 5 years

Inadequate response, loss of response, or intolerance to corticosteroid treatment is defined as one or more of the following:

Steroid refractory: persistent symptoms of active disease despite treatment with at least one 4-week induction regimen that included a dose of ≥ 30 mg prednisone (oral) daily (or equivalent) for at least 2 weeks or intravenously for at least 1 week within the previous 5 years

Steroid dependent: two failed attempts to taper steroids below a dose equivalent to 10 mg prednisone (oral) daily

Steroid intolerant: history of intolerance to corticosteroids (including, but not limited to, Cushing's syndrome, osteopenia/osteoporosis, hyperglycemia, insomnia, infection) within the previous 5 years

- Any ongoing UC therapy must be at stable doses:

  May be receiving oral 5-ASA compounds provided that the dose has been stable for ≥ 4 weeks immediately prior to Day 1

  May be receiving oral corticosteroid therapy (prednisone at a stable dose of ≤30 mg a day, or equivalent steroid).

  If corticosteroid therapy is ongoing or has recently been initiated, the dose has to be stable for at least 4 weeks immediately prior to Day 1. If corticosteroids are being tapered, the dose has to be stable for at least 2 weeks immediately prior to Day 1.

  May be receiving budesonide therapy at a stable dose of up to 9 mg a day for ≥ 4 weeks prior to Day 1

  May be receiving probiotics (e.g., Culturelle, Saccharomyces boulardii), provided that the dose has been stable for ≥ 2 weeks immediately prior to Day 1

  May be receiving AZA, 6-MP, or MTX, provided that the dose has been stable for ≥ 8 weeks immediately prior to Day 1

- For women who are not postmenopausal (at least 12 months of non−therapy-induced amenorrhea) or surgically sterile (e.g., absence of ovaries and/or uterus): agreement to remain abstinent or use a highly effective method of contraception during the treatment period and for at least 24 weeks after the last dose of study drug.

  The reliability of sexual abstinence should be evaluated in relation to the duration of the clinical trial and the preferred and usual lifestyle of the patient. Periodic abstinence (e.g., calendar, ovulation, symptothermal, or postovulation methods) and withdrawal are not acceptable methods of contraception.

- For men: agreement to remain abstinent (refrain from heterosexual intercourse) or use contraceptive measures, and agreement to refrain from donating sperm, as defined below:

  With female partners of childbearing potential or pregnant female partners, men must remain abstinent or use a condom during the treatment period and for at least 24 weeks after the last dose of study drug to avoid exposing the embryo to study drug. Men must refrain from donating sperm during this same period.

  The reliability of sexual abstinence should be evaluated in relation to the duration of the clinical trial and the preferred and usual lifestyle of the patient. Periodic abstinence (e.g., calendar, ovulation, symptothermal, or postovulation methods) and withdrawal are not acceptable methods of contraception.

- Must have received a colonoscopy within the past year or be willing to undergo a colonoscopy in lieu of a flexible sigmoidoscopy at screening. This colonoscopy must:

  Confirm disease extent (defined as 1) left-sided colitis [up to the splenic flexure], 2) extensive colitis [beyond the splenic flexure but not involving the entire colon], and 3) pancolitis; see the protocol)

  Include removal of any adenomatous polyps

  Document evidence of surveillance for dysplasia for all patients with left-sided colitis of • 12 years' duration and total/extensive colitis of • 8 years duration

Patients who meet any of the following criteria will be excluded from study entry:

Exclusion Criteria Related to Inflammatory Bowel Disease

- Prior extensive colonic resection, subtotal or total colectomy, or planned surgery for UC
- Past or present ileostomy or colostomy
- Diagnosis of indeterminate colitis
- Suspicion of ischemic colitis, radiation colitis, or microscopic colitis
- Diagnosis of toxic megacolon within 12 months of initial screening visit

- Any diagnosis of Crohn's disease
- Past or present fistula or abdominal abscess
- A history or current evidence of colonic mucosal dysplasia
- Patients with any stricture (stenosis) of the colon
- Patients with history or evidence of adenomatous colonic polyps that have not been removed

Exclusion Criteria Related to Prior or Concomitant Therapy

- Any prior treatment with etrolizumab or other anti-integrin agents (including natalizumab, vedolizumab, and efalizumab)
- Any prior treatment with anti-adhesion molecules (e.g., anti-MAdCAM-1)
- Any prior treatment with rituximab
- *Any treatment with tofacitinib during screening*
- Use of intravenous (IV) steroids within 30 days prior to screening with the exception of a single administration of IV steroid
- Use of agents that deplete B or T cells (e.g., alemtuzumab or visilizumab) within 12 months prior to Day 1, with the exception of AZA and 6-MP
- Use of cyclosporine, tacrolimus, sirolimus, or mycophenolate mofetil (MMF) within 4 weeks prior to Day 1
- Chronic nonsteroidal anti-inflammatory drug (NSAID) use (Note: occasional use of NSAIDs and acetaminophen [e.g., headache, arthritis, myalgias, or menstrual cramps] and aspirin up to 325 mg daily is permitted.)
- Patients who are currently using anticoagulants including, but not limited to, warfarin, heparin, enoxaparin, dabigatran, apixaban, rivaroxaban. (Note that antiplatelet agents such as aspirin up to 325 mg daily or clopidogrel are permitted.)
- Patients who have received treatment with corticosteroid enemas/suppositories and/or topical (rectal) 5-ASA preparations within 2 weeks prior to Day 1
- Apheresis (i.e., Adacolumn apheresis) within 2 weeks prior to Day 1
- Received any investigational treatment including investigational vaccines within 5 half-lives of the investigational product *or 28 days after the last dose*, whichever is greater, prior to Day 1
- History of moderate or severe allergic or anaphylactic/anaphylactoid reactions to chimeric, human, or humanized antibodies, fusion proteins, or murine proteins or hypersensitivity to etrolizumab (active drug substance) or any of the excipients (L-histidine, L-arginine, succinic acid, polysorbate 20)
- Patients administered tube feeding, defined formula diets, or parenteral alimentation/nutrition who have not discontinued these treatments $\geq$ 3 weeks prior to Day 1

Exclusion Criteria Related to General Safety

- Pregnant or lactating
- Lack of peripheral venous access
- Hospitalized (other than for elective reasons) during the screening period
- Inability to comply with study protocol, in the opinion of the investigator
- Significant uncontrolled comorbidity, such as cardiac (e.g., moderate to severe heart failure New York Heart Association [NYHA] Class III/IV), pulmonary, renal, hepatic, endocrine, or gastrointestinal disorders (excluding UC)
- Neurological conditions or diseases that may interfere with monitoring for PML
- History of demyelinating disease
- Clinically significant abnormalities on screening neurologic examination (PML Objective Checklist)
- Clinically significant abnormalities on the screening PML Subjective Checklist

- History of alcohol, drug, or chemical abuse ≤ 6 months prior to screening
- Conditions other than UC that could require treatment with • 10 mg/day of prednisone (or equivalent) during the course of the study
- History of cancer, including hematologic malignancy, solid tumors, and carcinoma in situ, within 5 years before screening with the following exceptions:

  Local basal or squamous cell carcinoma of the skin that has been excised and is considered cured is not exclusionary.

  A history of chronic myelogenous leukemia, hairy cell leukemia, melanoma, renal cell carcinoma, or Kaposi sarcoma is exclusionary, irrespective of the duration of time before screening.

  History of a cervical smear indicating the presence of adenocarcinoma in situ (AIS), high-grade squamous intraepithelial lesions (HSIL), or cervical intraepithelial neoplasia (CIN) of Grade • 1 is exclusionary, irrespective of the duration of time before screening.

Exclusion Criteria Related to Infection Risk

- Congenital or acquired immune deficiency
- Patients must undergo screening for HIV and test positive for preliminary and confirmatory tests
- Positive hepatitis C virus (HCV) antibody test result, unless the patient (1) has undetectable HCV RNA levels for >6 months after completing a successful course of HCV anti-viral treatment and an undetectable HCV RNA at screening or (2) has a known history of HCV antibody positivity with a history of undetectable HCV RNA and undetectable HCV RNA at screening in the absence of history of HCV anti-viral treatment.
- Patients must undergo screening for hepatitis B virus (HBV). This includes testing for HBsAg (HBV surface antigen), anti-HBc total (HBV core antibody total), and HBV DNA (patients who test negative for these tests are eligible for this study):

  Patients who test positive for surface antigen (HBsAg +) are not eligible for this study, regardless of the results of other hepatitis B tests.

  Patients who test positive only for core antibody (anti-HBc +) must undergo further testing for hepatitis B DNA (HBV DNA test).

    If the HBV DNA test is positive, the patient <u>is not eligible</u> for this study.

    In the event the HBV DNA test cannot be performed, the patient <u>is not eligible</u> for this study.

    If the HBV DNA test is negative, the patient <u>is eligible</u> for this study). These patients will undergo periodic monitoring for HBV DNA during the study.
- Evidence of or treatment for *Clostridium difficile* (as assessed by *C. difficile* toxin testing) within 60 days prior to Day 1 or other intestinal pathogens (as assessed by stool culture and ova and parasite evaluation) within 30 days prior to Day 1
- Evidence of or treatment for clinically significant cytomegalovirus (CMV) colitis (based on the investigator's judgment) within 60 days prior to Day 1. Laboratory confirmation of CMV from colon biopsy sample is required during screening evaluation only if clinical suspicion is high and to determine the need for CMV treatment
- History of active or latent treated tuberculosis (TB), regardless of treatment history (see protocol)

  Patients with a history of active or latent TB (based on a positive screening assay, either purified protein derivative [PPD] skin test or QuantiFERON→TB Gold test), are not eligible for this study.

  Patients with a chest X-ray (posteroanterior [PA] and lateral) within 3 months of Day 1 suspicious for pulmonary TB are not eligible for this study.
- History of recurrent opportunistic infections and/or history of severe disseminated viral infections (e.g., herpes)
- Any serious opportunistic infection within the last 6 months

- Any current or recent signs or symptoms (within 4 weeks before screening and during screening) of infection, except for the following:

  Minor infections (e.g., common cold) that have, in the investigator's judgment, completely resolved prior to Day 1

  Fungal infections of the nail beds

  Oral or vaginal candidiasis that has resolved with or without treatment prior to Day 1
- Any major episode of infection requiring treatment with IV antibiotics within 8 weeks prior to screening or oral antibiotics within 4 weeks prior to screening

  Treatment with antibiotics as adjunctive therapy for UC in the absence of documented infection is not exclusionary.
- Received a live attenuated vaccine within 4 weeks prior to Day 1
- History of organ transplant

Exclusion Criteria Related to Laboratory Values (at Screening)

- Serum creatinine • 2 times upper limit of normal (ULN)
- ALT or AST • 3 × ULN, or alkaline phosphatase • 3 × ULN, or total bilirubin • 2.5 × ULN (unconjugated hyperbilirubinemia that is associated with known Gilbert's syndrome is not an exclusion criterion)
- Platelet count • 100,000/μL
- Hemoglobin • 8 g/dL
- Absolute neutrophil count • 1500/μL
- Absolute lymphocyte count • 500/μL

Length of Study

The total length of the treatment period will be 62 weeks. Patients who do not achieve a clinical response at Week 10, patients who have clinical relapse during the Maintenance Phase, patients who receive defined rescue treatment (see protocol), and patients who complete 62 weeks of the study may be given the option of enrolling in Part 1 (OLE) of Study GA28951, where they will receive open-label etrolizumab treatment. Those who do not enroll in Part 1 (OLE) of Study GA28951 will continue to 12 weeks of safety follow-up in this study, and then be requested to enroll in Part 2 (SM) of Study GA28951 for 92 weeks of PML monitoring.

The total length of the study is expected to last from the first patient screened to either the last patient in last follow-up visit in this protocol or last patient enrolled into Study GA28951 (OLE-SM), whichever is later.

End of Study

The end of the study is defined as the last patient last safety follow-up visit in this protocol or last patient in this protocol enrolled into Study GA28951 (OLE-SM), whichever is later.

Efficacy Outcome Measures

The efficacy outcome measures for this study are as follows:

Primary Outcome Measure

- Remission at Week 62 among *patients with a clinical response* at Week 10*, as determined by the MCS*

Secondary Outcome Measures

- Clinical remission at Week 62 among patients in clinical remission at Week 10
- Clinical remission at Week 62
- *Remission at Week 62 among patients in remission at Week 10*
- Improvement in endoscopic appearance of the mucosa at Week 62
- Endoscopic remission at Week 62
- Histologic remission at Week 62
- *Change from baseline in rectal bleed subscore at Week 6*

- *Change from baseline in stool frequency subscore at Week 6*
- Corticosteroid-free clinical remission at Week 62 (off corticosteroids for at least 24 weeks prior to Week 62) in patients who are receiving corticosteroids at baseline
- Corticosteroid-free remission at Week 62 (off corticosteroids for at least 24 weeks prior to Week 62) in patients who are receiving corticosteroids at baseline
- Change from baseline to Week 62 in UC bowel movement signs and symptoms as assessed by the UC-PRO/SS measure
- Change from baseline to Week 62 in UC abdominal symptoms as assessed by the UC-PRO/SS measure
- Change from baseline to Week 62 in patients' health-related QOL as assessed by the overall score of the IBDQ

Exploratory Efficacy Outcome Measures

- *Clinical response at Week 62 among patients with a clinical response at Week 10*
- *Remission at Week 62 among patients in clinical remission at Week 10*
- Corticosteroid-free clinical remission at Week 62 (off corticosteroids for at least 12 weeks prior to Week 62) in patients who are receiving corticosteroids at baseline
- *Histologic disease activity change from baseline to Week 10 and Week 62*
- *Improvement in histologic and/or endoscopic disease activity*
- Change in health utilities, as assessed by the EQ-5D, from baseline to Weeks 10 and 62
- Improvement in endoscopic appearance of the mucosa at Week 10
- Endoscopic remission at Week 10
- Frequency and duration of hospitalizations from Week 10 to Week 62
- *Response, remission and corticosteroid-free endpoints as determined by the mMCS*

Safety Outcome Measures

The safety outcome measures for this study are as follows:

- Incidence and severity of adverse events
- Incidence of serious adverse events
- Incidence and severity of infection-related adverse events
- Incidence of serious infection-related adverse events
- Incidence and severity of injection-site reactions
- Incidence of adverse events leading to study drug discontinuation
- Incidence of laboratory abnormalities
- Incidence of malignancies
- Incidence of ATAs to etrolizumab
- Incidence and severity of hypersensitivity reaction events

Pharmacokinetic Outcome Measures

The PK outcome measures for this study are as follows:

- Serum trough concentration at steady state during the dosing period from Week 12 to Week 62
- Serum concentration at primary endpoint time (Week 62)

Exploratory Biomarker Outcome Measures

The exploratory biomarker outcome measures for this study are as follows:

- Remission at Week 10 and maintenance of remission at Week 62 in patients according to baseline levels of colonic tissue biomarkers and/or peripheral blood, including, but not limited to, αE integrin

- Changes in levels of exploratory colonic tissue and/or peripheral blood biomarkers during the Induction and Maintenance Phases
- Changes in stool biomarkers, which may include, but are not limited to, those in the microbiota and bacterial cultures, during the Induction and Maintenance Phases

Investigational Medicinal Products

Test Product

Etrolizumab prefilled syringe (PFS): containing SC formulation, 105 mg given as 0.7 mL of a 150-mg/mL solution will be administered by SC injection Q4W.

Comparator

Placebo PFS: etrolizumab SC matching placebo given in the amount of 0.7-mL solution will be administered by SC injection Q4W.

Non-Investigational Medicinal Products

Patients are to continue on their baseline dose of corticosteroid (including budesonide) to the end of the Induction Phase (Week 10). Tapering of corticosteroid (including budesonide) is to be attempted during the Maintenance Phase.

All patients are to continue on their baseline dose of immunosuppressants (AZA, 6-MP, MTX) to the end of Induction Phase (Week 10), unless dose reduction or discontinuation is required due to toxicity.

Probiotics and oral 5-ASA may be continued at a stable dose throughout the study.

Occasional use of NSAIDs and acetaminophen (e.g., headache, arthritis, myalgias, and menstrual cramps) and aspirin up to 325 mg daily are permitted throughout the study.

Antidiarrheals (e.g., loperamide, diphenoxylate with atropine) for control of chronic diarrhea are permitted throughout the study.

Statistical Methods

Primary Analysis

The analysis of data from the 62-week treatment period will be performed when all data from this period are in the database and data have been cleaned and verified.

Whereas Sponsor personnel will be unblinded to treatment assignment to perform the primary Week 62 analyses, patients and study site personnel will remain blinded to individual treatment assignment during the Maintenance Phase until after the study is completed (after all patients have either completed the safety follow-up period or discontinued early from the study) and the database is locked.

Detailed specifications of the statistical methods will be described in the Statistical Analysis Plan.

Determination of Sample Size

The study sample size was selected so that sufficient patients are enrolled to evaluate the primary endpoint in the Maintenance Phase. Approximately 350 patients will be enrolled in the open-label Induction Phase.

Under the assumption of *clinical response* at Week 10 of approximately *60%* (based on the Phase II Week 10 *response* results for the TNF-naive subgroup), the planned size of 350 patients in the open-label Induction Phase would provide approximately *210* patients *randomized into the Maintenance Phase.* A sample size of approximately *105* patients per arm is required in the Maintenance Phase to achieve*>90*% power to detect a 25% absolute difference in remission rates between the etrolizumab and placebo arms, for a χ^2 test at the two-sided 5% significance level, under the assumption of a placebo true remission rate of 30%−45%. If *clinical response* rates at Week 10 are lower than projected, the total enrollment in the open-label induction cohort may be increased to ensure that a *sufficient* number of Week-10 *clinical responders* are randomized into the Maintenance Phase.

Appendix 2
Schedule of Assessments

Assessments	Screening Day [a] −35 to −1	Study Week (±3 days)																		Unscheduled Visit [d]	Early Withdrawal from Treatment Visit
		0 [b]	4	8	10	12	16	20	24	28 [c]	32	36 [c]	40 [c]	44	48 [c]	52 [c]	56	60 [c]	62		
Informed consent	x																				
Review eligibility criteria	x	x [e]																			
Demographic data	x																				
Pregnancy test [f]	x	x [e]	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x		X
Vital signs (BP and pulse)	x	x [e]	x	x		x	x	x	x		x			x			x		x		X
ECG	x																		x		X
Chest X-ray [g]	x																				
Height		x																			
Weight		x																			
Medical history	x																				
Physical examination [h]	x					x		x			x			x			x		x		X
PML neurologic examination [i]	x		x			x		x			x			x			x		x	x [d]	X
Hematology	x	x [e]			x			x						x					x	x [d]	x [j]
Chemistry	x	x [e]			x			x						x					x	x [d]	x [j]
Urinalysis	x	x [e]																		x [d]	
TB screen [k]	x																				

Appendix 2
Schedule of Assessments (cont.)

		Study Week (±3 days)																			
Assessments	Screening Day [a] −35 to −1	0 [b]	4	8	10	12	16	20	24	28 [c]	32	36 [c]	40 [c]	44	48 [c]	52 [c]	56	60 [c]	62	Unscheduled Visit [d]	Early Withdrawal from Treatment Visit
HIV test	x																				
Hepatitis B and C serology [l]	x																				
Hepatitis B DNA [m]	x					x			x		x			x					x		
Hepatitis C RNA (Amplicor) [n]	x																				
Flow cytometry		x			x														x		
PK sampling (serum) [o]		x				x			x					x					x	x [d]	x [j]
Anti-therapeutic antibody sample (serum) [o, p]		x	x			x			x					x					x [q]	x [d]	x [j, q]
Plasma sample [r] (storage for JCV antibody testing)	x																				
MCS (includes endoscopy) [s, t]	x [u]				x														x	x [d]	x [j]
pMCS (excludes endoscopy) [v]		x [u]	x	x		x	x	x	x		x			x			x			x [d]	x [j]
Stool sample	x [w]	x [x]			x [x]														x [x]	x [d]	x [x]
Colonic biopsy (CMV if required)	x [y]																			x [d]	

Appendix 2
Schedule of Assessments (cont.)

		Study Week (±3 days)																			
Assessments	Screening Day [a] −35 to −1	0 [b]	4	8	10	12	16	20	24	28 [c]	32	36 [c]	40 [c]	44	48 [c]	52 [c]	56	60 [c]	62	Unscheduled Visit [d]	Early Withdrawal from Treatment Visit
Colonic biopsy (histopathological confirmation of UC if required)	x [z]																				
Colonic biopsy (formalin)	x [aa]				x [bb]														x [bb]	x [d, bb]	x [j, bb]
Colonic biopsy (RNA later)	x [aa]				x [bb]														x [bb]	x [d, bb]	x [j, bb]
Serum sample (CRP)		x [e]			x														x	x [d]	x [j]
Blood sample for genetic analysis (DNA) (optional)		x																			
Serum sample (future exploratory PD) [o]		x				x			x					x					x		X
Blood sample (RNA Paxgene) [o, cc]		x			x				x										x		X
UC-PRO/SS measure [t]		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
IBDQ [dd]		x			x														x		
EQ-5D [dd]		x			x						x								x		
Concomitant medications		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Appendix 2
Schedule of Assessments (cont.)

		Study Week (±3 days)																			
Assessments	Screening Day [a] −35 to −1	0 [b]	4	8	10	12	16	20	24	28 [c]	32	36 [c]	40 [c]	44	48 [c]	52 [c]	56	60 [c]	62	Unscheduled Visit [d]	Early Withdrawal from Treatment Visit
Adverse events	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Randomization of responders to Maintenance Phase					x [ee]																
Etrolizumab administration		x	x	x																	
Etrolizumab/ etrolizumab placebo administration [ff]						x	x	x	x	x	x	x	x	x	x	x	x	x			

ATA = anti-therapeutic antibody; BP = blood pressure; CMV = cytomegalovirus; CRP = C-reactive protein; EQ-5D = EuroQoL Five-Dimension Questionnaire; HBc = HBV core antibody total; HBsAg = HBV surface antigen; HBV = hepatitis B virus; HCV = hepatitis C virus; IBDQ = Inflammatory Bowel Disease Questionnaire; JCV = John Cunningham virus; MCS = Mayo Clinic Score; PD = pharmacodynamics; PK = pharmacokinetic; pMCS = partial Mayo Clinic Score; PML = progressive multifocal leukoencephalopathy; qPCR = quantitative polymerase chain reaction; TB = tuberculosis; tc = telephone call; UC-PRO/SS = Ulcerative Colitis Patient-Reported Outcome Signs and Symptoms.

Notes: All study assessments and blood draws are to be conducted prior to study drug administration. All colonic biopsy samples will be taken during flexible sigmoidoscopy/colonoscopy procedure.

[a] All assessments must be performed after obtaining informed consent. Endoscopy should be performed 4–16 days prior to Day 1 (i.e., Day −16 to Day −4). The total screening period is 35 days. Under no circumstances will either window be extended.

[b] Day 1 of Week 0.

[c] Telephone contact for patients performing home administration: Patients requiring in-clinic drug administration throughout the study will have their study assessments conducted in clinic or via telephone call after their clinic visit at the sites' discretion.

Appendix 2
Schedule of Assessments (cont.)

[d] Unscheduled visit represents a visit that is not per Schedule of Assessments and is required for an adverse event or for potential relapse assessment. All indicated assessments are NOT performed at each unscheduled visit. Assessments would be symptom driven (e.g., only perform PML neurologic examination if patient reports symptoms suspected of PML; for disease worsening, infectious etiologies may be investigated if clinically indicated; and confirmation of clinical relapse is performed by the MCS assessment). Assessments corresponding to items noted in this column should be recorded on the eCRF.

[e] Perform prior to administration of etrolizumab.

[f] Serum test at screening for all female patients except those who are more than 1 year postmenopausal or have had a hysterectomy. Urine test at other visits; if urine test result is positive, perform a confirmatory serum test. Pregnancy test will be carried out at home once patient starts etrolizumab administration at home. Patient is to report the pregnancy test via the e-diary. Patients must be instructed at screening and reminded throughout the study that in case of positive pregnancy test they must stop self-administration of study drug and call the site immediately. Do not administer etrolizumab unless the serum pregnancy test result is negative.

[g] Not required if normal chest X–ray result within 3 months prior to screening.

[h] Full physical examination required at screening; symptom-driven physical examination at all other timepoints indicated.

[i] PML neurologic examination consists of the PML Subjective Checklist and the PML Objective Checklist. Administer before other assessments, per Appendix 5.

[j] Not required if unscheduled visit leads to withdrawal and assessment previously conducted at unscheduled visit.

[k] The following tests are acceptable screening assays for latent TB in this study: purified protein derivative (a tuberculin skin test reaction; e.g., Mantoux test), INF-γ based test (e.g., QuantiFERON®-TB Gold).

[l] Patients must undergo screening for HBV and HCV. This includes testing for HBsAg, anti-HBc, and hepatitis C antibody.

[m] Enrolled patients who are hepatitis B core antibody positive should have hepatitis B DNA measured at these timepoints.

[n] Measurement of HCV RNA with use of the Amplicor assay is required when the patient has a known history of HCV antibody positivity with past documentation of undetectable HCV RNA, either with or without history of anti-viral treatment. Patients with newly diagnosed HCV antibody positivity are not eligible for this study and, therefore, do not require measurement of HCV RNA.

[o] All samples to be collected prior to administration of etrolizumab. Serum baseline, Week 4, and blood sample (RNA Paxgene) are to be collected prior to dose of etrolizumab from all patients. All subsequent samples (serum and Paxgene whole blood) will be collected only from those who were eligible to be randomized into the Maintenance Phase.

[p] If serum sickness or a clinically significant allergic drug reaction is suspected, Sponsor should be notified, and serum for etrolizumab level and ATAs should be drawn and sent to the central laboratory. ATA samples may also be utilized for exploratory PD assessments or assessment of drug concentrations.

Appendix 2
Schedule of Assessments (cont.)

[q] Collection of sample for ATA is required at final or early withdrawal visit, unless it coincides with first visit in Part 1 of Study GA28951 (where a sample for ATA must be collected).

[r] A blood sample to test for antibodies to JCV will be taken and stored for possible later assessment of how widespread the JCV infection is in the study population. Sample testing for the presence of JCV antibodies is not helpful in predicting risk for PML or for evaluating neurologic symptoms. The sample may be tested if there is a strong belief that this information will be helpful in managing a patient's condition.

[s] Endoscopy + rectal bleeding assessment + stool frequency assessment + Physician's Global Assessment. Patients who have not undergone full colonoscopy with documented results within 1 year prior to screening should undergo colonoscopy in lieu of sigmoidoscopy at the screening visit to allow for screening for cancer/dysplasia (yes/no).

[t] During screening, patients must be trained on the use of the e-diary. Patients are to complete the e-diary on a daily basis for the stool frequency and rectal bleeding score (for MCS/pMCS), starting from the first screening visit, and for at least 9 – 12 consecutive days around the time of each scheduled visit for the UC-PRO/SS.

[u] Screening endoscopy (for the MCS) should be performed 4 – 16 days prior to Day 1 (i.e., Day –16 to Day – 4). For baseline measurements, the Physician's Global Assessment will be obtained only once, on Day 1 (prior to enrollment), and the Physician's Global Assessment score will be used to calculate both the baseline (screening) MCS and the baseline (Day 1) pMCS.

[v] Rectal bleeding assessment + stool frequency assessment + Physician's Global Assessment.

[w] For culture and sensitivity testing; ova, parasites, and Clostridium difficile toxin testing.

[x] Sample analyses may include, but are not limited to, analyses of fecal calprotectin and other exploratory PD biomarkers (such as analyses of the microbiota and bacterial cultures).

[y] IF REQUIRED: Only if there is suspicion for clinically significant CMV colitis, one biopsy sample should be obtained from the base of the ulcer to evaluate for histological presence of CMV. Analysis should be performed locally if possible, or can be sent to a central laboratory if necessary. Result must be negative for CMV prior to dosing on Day 1.

[z] IF REQUIRED: If patient does not have previously documented histopathologic confirmation of UC as defined in the inclusion criteria, one biopsy sample can be obtained from the base of the ulcer read locally for histopathologic confirmation of UC.

[aa] In addition to the optional biopsy samples noted in footnote "y" and "z" above, five pairs (10 biopsy samples) will be obtained at screening. These five biopsy pairs will be sent to the central laboratory for further storage or distribution. Two pairs (taken from the most inflamed area of colon within 20−40 cm of anal verge [sigmoid]) will be placed in stabilization buffer (such as RNAlater or a similar buffer) and stored at −80°C (one pair for diagnostic qPCR and one pair for PD biomarkers qPCR). The other three pairs (two pairs from most inflamed area of colon within 20−40 cm of anal verge [sigmoid] and one pair from the most inflamed area of the worst affected segment) will be placed in formalin and then paraffin embedded (one pair will be used for histopathology and exploratory PD biomarkers and the other two pairs will be used for diagnostic). Original biopsy location and endoscopic depth should be clearly indicated.

Appendix 2
Schedule of Assessments (cont.)

[bb] A total of four pairs (8 biopsy samples) will be obtained. All will be sent to the central laboratory for further storage or distribution. One pair will be placed in stabilization buffer (such as RNAlater or a similar buffer) and stored at −80°C for exploratory PD or diagnostic biomarker qPCR. The other three pairs, representing three different segments (rectum, sigmoid, descending colon), will be placed in formalin and then paraffin embedded; these biopsy samples will be used for histopathology, exploratory PD biomarkers, and/or diagnostic biomarker. Original biopsy location and endoscopic depth should be clearly indicated.

[cc] Paxgene blood RNA samples must be collected after all other blood and serum samples.

[dd] With the exception of Week 0, the IBDQ and the EQ-5D will be completed in-clinic by the patient after the PML neurological examination but before any other non-PRO assessments and before the patient receives any disease status information or study drug during that visit.

[ee] Randomization to occur within 2 weeks starting from Week 10 timepoint.

[ff] Where indicated, patients must be instructed to administer study drug at home within 3 days (maximum) after clinic visit.

Appendix 3 Rescue Therapy

Study Phase	Type	Description
Induction	Permitted	Initiation or escalation beyond baseline dose of the following agents for the treatment of worsening UC symptoms. • Immunosuppressants (AZA, 6-MP or MTX) • Oral or topical 5-ASA or corticosteroid (use of topical at baseline is among the exclusion criteria; rectal is synonymous with topical)
	Prohibited	Any use of other immunosuppressants, including: • TNF inhibitors and biosimilars thereof • Cyclosporine, tacrolimus, sirolimus, or MMF • Anti-adhesion molecules, including natalizumab and vedolizumab • Other biologics, such as efalizumab, alemtuzumab, visilizumab and rituximab • Other investigational agents, including vaccines (e.g., MAP or ChAdOx2 HAV, JAK inhibitors))
Maintenance	Permitted	Same definition as induction, noting that: • Maintenance period begins with an oral corticosteroid taper • All patients receiving rescue therapy during the induction period are ineligible to enter maintenance
	Prohibited	Same definition as induction, noting that: • All patients receiving rescue therapy during the induction period should withdraw from the study

5-ASA=5-aminosalicylic acid; 6-MP= 6-mercaptopurine; AZA=azathioprine; ChAdOx2 HAV=chimpanzee adenovirus Oxford 2 hepatitis A vaccine; JAK= Janus kinase; MAP=Mycobacterium avium subspecies paratuberculosis; MMF=mycophenolate mofetil; MTX=methotrexate; TNF=tumor necrosis factor; UC=ulcerative colitis.

STATISTICAL ANALYSIS PLAN

TITLE: **DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTICENTER STUDIES OF THE EFFICACY AND SAFETY OF ETROLIZUMAB DURING INDUCTION AND/OR MAINTENANCE PHASES IN PATIENTS WITH MODERATE TO SEVERE ACTIVE ULCERATIVE COLITIS WHO HAVE BEEN PREVIOUSLY EXPOSED TO TNF INHIBITORS OR PATIENTS WHO ARE aTNF NAÏVE**

PROTOCOL NUMBER(S): RO5490261 (GA28948, GA28949, GA29102, GA29103, GA28950)

STUDY DRUG: Etrolizumab

VERSION NUMBER: 3

IND NUMBER: 100366

EUDRACT NUMBER: 2013-004278-88, 2013-004280-31, 2013-004279-11, 2013 004277-27, 2013-004282-14

SPONSOR: F. Hoffmann-La Roche Ltd.

PLAN PREPARED BY: [redacted], M.Sc.

DATE FINAL: Version 1: 14 May 2019

DATE AMENDED: Version 2: 17 December 2019
Version 3: See electronic date stamp below.

STATISTICAL ANALYSIS PLAN AMENDMENT APPROVAL

Date and Time(UTC)	Reason for Signing	Name
04-May-2020 14:40:20	Company Signatory	[redacted]

STATISTICAL ANALYSIS PLAN AMENDMENT
RATIONALE FOR VERSION 3

- Changes to Section 4.3.2 (Baseline Disease Characteristics):
  - Prior tumor necrosis factor (TNF) Category updated Naive to Unknown
- Changes in Section 4.4.2.1.1 (Stool Frequency and Rectal Bleeding):
  - For analysis at baseline Mayo Clinic Score (MCS) subscores, stool frequency (SF) and rectal bleeding (RB) calculations clarified calculation when endoscopy occurred outside of window.
- Changes in Sections 4.4.2.1.2 (Endoscopy) and 4.4.2.1.4 (Outcome Measures Derived from the MCS):
  - Replaced using the terminology of overall score with the Hybrid Score (HS)
  - Add definition Endoscopic Remission (excluding friability from mild subscore 1)
  - Add definition Improvement of the endoscopic mucosa (excluding friability from mild subscore 1)
- Changes to Section 4.4.2.3 (UC-PRO):
  - PRO analysis method of analysis of covariance (ANCOVA) removed and mixed-model repeated measures (MMRM) has been added as the primary method.
  - Clarified the bowel domain score is 0-27
- Changes to Section 4.3 (Subgroups):
  - Added in anti-drug antibodies (ADA) positive Subgroups
- Changes to Section 4.7.1 (Adverse Events):
  - Added in non-treatment emergent report
  - Added rates for Studies GA28948 and GA28949
- Changes to Section 4.8 (Missing Data):
  - For continuous endpoints worst post baseline imputation has been removed for ulcerative colitis patient-reported outcome (UC-PRO) data, and observed case analysis added as a sensitivity analysis for SF/RB and Inflammatory Bowel Disease Questionnaire (IBDQ).
  - Further clarification of when non-responder imputation will be applied to categorical endpoints has been added.
  - Clarification the tipping point analyses will be conducted for all primary endpoints across all studies.

Additional minor changes have been made to improve clarity and consistency.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

GLOSSARY OF ABBREVIATIONS

AE	adverse event
AESIs	adverse events of special interest
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
aTNF	anti-tumor necrosis factor
AST	aspartate aminotransferase
CMH	Cochran-Mantel-Haenszel
CRP	C-reactive protein
CRF	Case Report Form
CS	corticosteroid
CSR	Clinical Study Report
ES	Endoscopic Subscore
FFPE	formalin-fixed paraffin-embedded
HS	Hybrid Sigmoid
IBD	inflammatory bowel disease
IBDQ	Inflammatory Bowel Disease Questionnaire
iDCC	independent Data Coordinating Center
iDMC	independent Data Monitoring Committee
IS	immunosuppressant
IxRS	interactive voice/web based response system
LoPO	list of planned outputs
MCS	Mayo Clinic Score
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified intent to treat
mMCS	modified Mayo Clinic Score
MMRM	mixed-model repeated measures
NHI	Nancy Histology Index
OLE	open label extension
OLI	open label induction
PD	pharmacodynamic
PGA	Physician’s Global Assessment
PK	pharmacokinetic
pMCS	partial Mayo Clinic Score
RB	rectal bleeding
RHI	Robarts Histopathological Index
SAE	serious adverse event
SAP	Statistical Analysis Plan
SF	stool frequency
SMQ	standardized MedDRA query
TNF	tumor necrosis factor
UC	ulcerative colitis
UC-PRO/SS	Ulcerative Colitis Patient-Reported Outcome Signs and Symptoms
ULN	upper limit of normal

GLOSSARY OF ABBREVIATIONS

WOCF Worse Observation Carried Forward

1. BACKGROUND

This etrolizumab project Statistical Analysis Plan (SAP) describes the study design and analyses that are common to the Phase III etrolizumab studies in ulcerative colitis (UC) patients detailed in Table 1. Elements of the study design and analysis unique to the individual studies will be explained in the respective study SAPs.

Table 1 Study Descriptions

Study	Study Description
GA28948 (Hibiscus I) and GA28949 (Hibiscus II)	Two identical placebo-controlled induction studies assessing the efficacy and safety of etrolizumab compared to adalimumab and placebo in patients with moderate to severe active ulcerative colitis who are naive to TNF inhibitors
GA29102 (Laurel)	Placebo-controlled, maintenance study assessing the efficacy and safety of etrolizumab in patients with moderate to severe active ulcerative colitis who are naive to TNF inhibitors
GA29103 (Gardenia)	Head To Head study to evaluate the efficacy and safety of etrolizumab compared with Infliximab in patients with moderate to severe active ulcerative colitis who are naive to TNF inhibitors
GA28950 (Hickory)	Placebo-controlled, induction and maintenance study assessing the efficacy and safety of etrolizumab in patients with moderate to severe active ulcerative colitis who have been previously exposed to TNF inhibitors

TNF=tumor necrosis factor.

Studies GA28950 and GA29102 are made up of two phases: Induction Phase and Maintenance Phase. For the purpose of statistical analyses, the Induction and Maintenance Phases of Studies GA28950 and GA29102 will be treated as two independent studies and analyzed separately. Studies GA28948 and GA28949 are induction studies. Study GA29103 is a treat-through study design with no re-randomization into the Maintenance Phase; therefore, the Induction and Maintenance Phases will be analyzed as one study.

The analysis of data will be performed once all the data have been collected in the database for the primary analysis as described in the study SAPs.

2. STUDY DESIGN

2.1 PROTOCOL SYNOPSIS

For individual study Protocol Synopses and Schedules of Assessments, refer to the study SAP Appendix 1 and Appendix 2, respectively.

2.2 OUTCOME MEASURES

For individual study outcome measures, refer to the study SAPs.

2.3 DETERMINATION OF SAMPLE SIZE

For details of individual study sample size, refer to the study SAPs.

2.4 ANALYSIS TIMING

For details of the analysis timing for individual studies, refer to the study SAPs.

3. STUDY CONDUCT

3.1 RANDOMIZATION

An independent interactive voice/Web-based response system (IxRS) vendor will conduct the randomization for all studies and the independent Data Coordinating Center (iDCC) will perform regular checks of the randomization scheme using unblinded data. The patient randomization list will be generated by the IxRS with use of a pre-defined randomization specification. During study conduct the randomization list will not be available to the study sites, study monitors, project statisticians, or the Sponsor's project team. The study team will remain blinded to study drug. If unblinding is necessary for patient management (in the case of a serious adverse event [SAE]), the investigator will be able to break the treatment code by contacting the IxRS. Treatment codes should not be broken except in emergency situations. If the investigator wishes to know the identity of the study drug for any other reason, he or she should contact the Medical Monitor. The investigator should document and provide an explanation for any premature unblinding (e.g., accidental unblinding, unblinding due to an SAE as per health authority reporting requirements). The Sponsor safety reporting department (independent to the study team) will unblind the identity of the study medication for all unexpected SAEs that are considered by the investigator to be related to study drug per safety reference document(s), such as the Investigator's Brochure, Core Data Sheet, and Summary of Product Characteristics (SmPC). Details of patients who are unblinded during the study will be included in the individual study Clinical Study Reports (CSRs).

After the end of randomization, the data entered into the IxRS system will be reconciled with the data entered into the clinical database. In particular, the kit assignments and stratification factors will be checked. Discrepancies between the IxRS and clinical database will be listed in the CSR and raised as protocol deviations. The statistical analyses will be conducted using IxRS stratification factors, and sensitivity analysis using clinical database data will be conducted if required.

Further details of stratification factors for each study is included in the study SAPs.

3.2 INDEPENDENT REVIEW FACILITY

The efficacy measure Mayo Clinic Score (MCS) requires endoscopic subscores to be collected. Central reading of endoscopies will be performed throughout the studies by

an independent review facility. Data collected at sites as video clips are read centrally by an independent gastroenterologist experienced in inflammatory bowel disease (IBD). The independent reader will be blinded to the patient's clinical activity and treatment allocation. Reads are collected and read locally and the again centrally. The adjudication is carried out in two stages:

1. A second central reading is performed if the local and initial central reading do not agree or the initial central reading cannot be performed.
2. The local and central reading results are combined in a final Mayo Endoscopy subscore using the median among readers, rounded up to the nearest integer.

Further details are available in the Independent Review Charter.

The efficacy measure of histologic remission requires independent scoring. The histologic scoring will be performed by a small pool of central readers who are blinded both to treatment arm and timepoint. The scoring database will ensure that all slides for a given patient are scored by the same reader. Slide image scores are based on formalin-fixed paraffin-embedded (FFPE) biopsies from the most inflamed region of the sigmoid colon. Scores are queried for discrepancies between Nancy Histologic Index (NHI) and Geboes results (e.g., NHI < 4, indicating no erosions/ulcerations, and Geboes subgrades 5.3 or 5.4, indicating the presence of erosions/ulcerations). Queries may lead to the same reader reassessing the relevant slide images and revising the scores as they deem necessary. Further details are available in the Image Review Charter.

3.3 DATA MONITORING

An independent Data Monitoring Committee (iDMC) will monitor safety and study conduct on an ongoing basis for all studies. The iDMC will meet approximately every 6 months. Further details are available in the iDMC Charter.

4. STATISTICAL METHODS

4.1 ANALYSIS POPULATIONS

Analysis populations are reported in the individual study SAPs.

4.2 ANALYSIS OF STUDY CONDUCT

For details on the analysis of study conduct, refer to the individual study SAPs.

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

To review treatment group comparability within each study a number of variables collected at baseline will be compared across treatment groups. Baseline is defined as the last available assessment prior to first receipt of study drug.

For continuous variables, descriptive statistics including n, mean, median, SD, minimum, and maximum will be calculated. For categorical variables, number and percentage in

each category will be displayed. The units/categories to be used are indicated within the brackets and separated by commas.

Summaries by treatment group will be presented for all analysis populations. Demographics and baseline characteristics presented for patients in the Maintenance Phases will use the data collected from their baseline visits.

4.3.1 Demographics

Demographics presented will include;

- Age at randomization (years), descriptive statistics, and number and percentage of patients in the following categories: 18–<40, ≥40–<65, ≥65
- Gender, number and percentage of patients in the following categories: male, female
- Race, number and percentage of patients in the following categories:
  - American Indian or Alaskan Native
  - Asian
  - Black or African American
  - Native Hawaiian or other Pacific Islander
  - White
  - Other (includes Other, Multiple, and Unknown)

For Listings the Race Category Other, multiple and unknown will be reported. For summaries and subgroups the combined category 'Other' will be used.

- Region, number and percentage of patients in the following categories:
  - Central /Eastern Europe
  - USA
  - Western/Northern Europe, Canada, Australia, New Zealand
  - Asia
  - Latin America
  - Other
- Ethnicity, number and percentage of patients in the following categories: Hispanic or Latino, Not Hispanic or Latino, Not reported, Unknown
- Body Weight (kg), descriptive statistics
- Body Mass Index (BMI; kg/m^2), descriptive statistics
- Tobacco Use, number and percentage of patients in the following categories: Never, Previous, Current

4.3.2 Baseline Disease Characteristics

Details of efficacy parameters definitions listed below are provided in Section 4.4.2.

- Duration of disease (years), descriptive statistics, and number and percentage of patients in the following categories: <3, ≥3-<8, ≥8
- Disease extent, number and percentage of patients in the following categories: Left-sided colitis, Extensive colitis, Pancolitis
- Serum C-reactive protein (CRP) (mg/L), descriptive statistics, and number and percentage of patients in the following categories: ≤2.87, >2.87– ≤10, >10
- Fecal calprotectin (μg/g), descriptive statistics including median, 25th and 75th percentiles and number and percentage of patients in the following categories: <250, ≥250–<500, ≥500
- Mayo Clinic Score (MCS), descriptive statistics, and number and percentage of patients in the following categories: MCS ≤ 9, MCS ≥ 10
- Partial Mayo Clinic Score (pMCS) –Descriptive statistics
- Modified Mayo Clinic Score (mMCS) –Descriptive statistics
- Modified Mayo Clinic Score (mMCS) (excluding friability from Mayo ES=1) – Descriptive statistics
- Stratification Factor MCS Score per IxRS, number and percentage of patients in each category: MCS ≤ 9, MCS ≥ 10
- Stool Frequency (SF) – Descriptive statistics.
- Rectal Bleeding (RB) – Descriptive statistics.
- Physician's Global Assessment (PGA) – Descriptive statistics
- Endoscopy (ES) – Descriptive statistics.
- Ulcerative Colitis Patient-Reported Outcomes, Signs and Symptoms (UC-PRO/SS): Functional – Descriptive statistics
- UC-PRO/SS: Bowel – Descriptive statistics
- UC-PRO: Systemic Symptoms – Descriptive Statistics
- Inflammatory Bowel Disease Questionnaire (IBDQ) – Descriptive statistics
- Nancy Histological Index (NHI) – Descriptive statistics and frequencies (number and percentage per grade)
- Robarts Histopathological Index (RHI) – Descriptive statistics and frequencies (number and percentage per grade)
- Geboes Grading Scale Score – Descriptive statistics and frequencies (number and percentage per grade)

4.3.3 Baseline Disease Medications

All data types are categorical and the number and percentage of patients will be presented for each category denoted in brackets.

- Corticosteroid (CS) use at baseline (yes, no)
- Stratification Factors (IxRS): Corticosteroid (CS) use at baseline (yes, no)
- Immunosuppressant (IS) use at baseline (yes, no)
- Stratification Factors (IxRS): Immunosuppressant (IS) use at baseline number (yes, no)
- Corticosteroid and Immunosuppressant categories (CS alone, IS alone, CS and IS, None)
- Prior anti-tumor necrosis factor (aTNF) medication (yes, no)
- Prior aTNF medication (1 failure, ≥2 failures, Refractory =Primary Non-response, Loss of response =Secondary Loss of Response, Intolerant, Unknown)

Prior aTNF use is only relevant for Study GA28950; all patients in other studies are aTNF-naive.

4.4 EFFICACY ANALYSIS

The hierarchical priority of key secondary endpoints and analysis populations are available in the study SAPs. All formal statistical comparisons for binary data will use the Cochran-Mantel-Haenszel (CMH) test statistics stratified by the factors used at randomization as described in the study SAPs. For all analyses in both the Induction and Maintenance Phases where comparisons back to baseline are made, baseline is defined as the last available assessment prior to first receipt of study drug in the Induction Phase. For all analyses the point estimate, 95% CIs and p-value will be reported.

4.4.1 Efficacy Endpoints

All primary endpoints for the studies are derived from the MCS. Further details of the individual primary endpoint evaluation and treatment effects (estimands) are provided in the study SAPs. All primary endpoints are categorical and formal statistical comparisons between the treatment arms will use the CMH test statistics stratified by the factors used at randomization. In addition, a selection of secondary endpoints across the studies evaluating remission, clinical remission and clinical response are also derived using the full MCS.

4.4.2 Endpoint Definitions

This section provides endpoint definitions for endpoints common across studies. Further study specific details including analysis timepoints and populations will be included in the study SAPs.

4.4.2.1 Mayo Clinic Score

The MCS is a composite endpoint made up of four components. The score ranges from 0 to 12 with higher scores indicating more severe disease.

The MCS is used to determine a number of efficacy endpoints as described in Table 3.

$$\text{MCS} = \text{Stool Frequency subscore} + \text{Rectal Bleeding subscore} + \text{Endoscopy subscore} + \text{PGA subscore}$$

4.4.2.1.1 Stool Frequency and Rectal Bleeding

Stool frequency (SF) and rectal bleeding (RB) data are collected daily via patient's diaries and each day a patient provides a score between 0-3 for each component.

Stool Frequency Subscore 0–3

- 0 = Normal number of stools for this patient
- 1 = 1 to 2 stools more than normal
- 2 = 3 to 4 more stools than normal
- 3 = 5 or more stools than normal

Rectal Bleeding Subscore 0–3

- 0 = No blood in stool
- 1 = Streaks of blood with stool less than half the time
- 2=Obvious blood with stool most of the time
- 3 = Blood alone passed

– Stool Frequency (SF) subscore = Average of 3 days daily diary scores

– Rectal Bleeding (RB) subscore = Worst value of 3 days daily diary scores

Table 2 summarizes the different scenarios for calculating SF/RB sub score in the analyses. The three days of daily diary data used to calculate the subscores described above are selected from the days most recent to, (but not including) a pre specified date, and this is called the 'anchor date'.

Table 2 Steps for Calculating SF/RB Subscore

Timepoint	Scenarios to calculate SF/RB Subscore
Baseline	Scenario 1 Bowel preparation date (prior to an endoscopy) is assigned as the 'anchor' date. Three days daily diary data collected from patient's e-diary between Day -22 and the day prior to the anchor date will be selected to calculate SF/RB scores. Daily dairy days selected closest to the anchor date will be selected first
	Scenario 2 Bowel preparation date (prior to an endoscopy) is assigned as the 'anchor' date. If fewer than three days of daily diary data are available between Day −22 through to the day prior to the anchor date, then additional daily diary data collected post endoscopy starting with the score recorded 2 days after the endoscopy but prior to the randomization/enrollment will be selected to calculate SF/RB scores. Examples of this scenario are illustrated (see Appendix 3).
	Scenario 3 Randomization/Enrollment Date is assigned as the 'anchor' date, if endoscopy did not occur between Day -21 and day prior to randomization/enrollment date. Three days worth of diary data prior to the 'anchor' date will be selected to calculate SF/RB scores. Daily dairy days selected closest to the anchor date will be selected first
Post Baseline	Scenario 1 Bowel preparation date (prior to an endoscopy) occurring within 7 days prior to the visit date is assigned as the 'anchor' date. Three daily diary days collected within the 7 days prior to post baseline visit date will be selected, to calculate SF/RB scores. Daily dairy days selected closest to the anchor date will be selected first. Visit date is the day the PGA assessment is collected. Scenario 2 Bowel preparation date (prior to an endoscopy) is assigned as the 'anchor' date. Three daily diary days collected within the 7 days prior to post baseline visit date will be selected, to calculate SF/RB scores. Daily dairy days selected closest to the anchor date will be selected first. If further days are required, data collected +2 days after endoscopy within the 7 days prior to visit date will be used. Visit date is the day the PGA assessment is collected.
	Scenario 3 Visit date is assigned as the 'anchor' date. Three daily diary days collected within the 7 days prior to post baseline visit date will be selected, to calculate SF/RB scores. Daily dairy days selected closest to the anchor date will be selected first.

SF= stool frequency; RB= rectal bleeding; PGA= Physician's Global Assessment.

The change from baseline in the SF subscore and RB subscore at Week 6 between the treatment arms will be reported for GA28950, GA28949 and GA28948 studies. This data is considered non-parametric and will be reported using RANK analysis of covariance (ANCOVAs).

4.4.2.1.2 Endoscopy

Endoscopy = Assessment of segments from 3 locations (Colon Descending, Colon Sigmoid, Rectum).

Each location is scored using the following criteria:

Endoscopic Subscore 0–3

- 0 = Normal or inactive disease
- 1 = Mild disease (erythema, decreased vascular pattern, mild friability)
- 2 = Moderate disease (marked erythema, lack of vascular pattern, friability, erosions)
- 3 = Severe disease (spontaneous bleeding, ulceration)

At baseline all segments are reviewed and scored and the worst score from the three segments is recorded as the endoscopy subscore. At baseline the endoscopy score is collected – 16 to – 4 days prior to Day 1, however for analyses purposes, any endoscopy collected prior to randomization/enrollment date will be used as the baseline endoscopy score. All assessments are performed via video and assessed by both a local reader and a central reader; adjudication process is applied if required (see Section 3.2).

Post baseline the endoscopic score is the worst score of all segments which have been assessed at baseline, if the baseline endoscopy score had Sigmoid colon score ≤1. If at baseline the sigmoid colon score was ≥2, the post baseline endoscopy score is the sigmoid colon score value. This methodology is called the Hybrid Sigmoid (HS) model and is used as the primary method for all analyses across all studies and throughout the project and study SAP documents the HS methodology will be used when calculating endoscopic subscore.

For mMCS (excluding friability from ES = 1) the patient's endoscopic subscore will be updated from ES = 1 (mild disease) to ES = 2 (moderate disease) if friability is present for either central reader 1 or central reader 2 at any location. All other patients' endoscopic subscores will remain the same. For all other analyses in Table 3 using the endoscopic subscore, the definition with mild friability is considered within endoscopic subscore of 1, will be used.

4.4.2.1.3 Physician's Global Assessment

The PGA will be provided by the investigator as a score of 0 to 3. The status is based on the physician's overall rating of the patient's disease activity, given endoscopy, stool frequency, rectal bleeding, abdominal pain, well-being, fecal continence, observations, and physical exam findings.

Physician's Global Assessment

- 0 = Normal (Subscores are 0)
- 1 = Mild disease (Subscores are mostly 1s)
- 2 = Moderate disease (Subscores are 1 to 2)
- 3 = Severe disease (Subscores are 2 to 3)

4.4.2.1.4 Outcome Measures Derived from the MCS

The outcome measures calculated using the MCS or a selection of components from the MCS are detailed in Table 3.

Table 3 Outcome Measures

Outcome Measure	Definition
Mayo Clinic Score (MCS)	MCS (0–12) is a composite of 4 assessments, each rated from 0–3: stool frequency, rectal bleeding, endoscopy, and PGA
Partial Mayo Clinic Score (pMCS)	pMCS (0–9) is a composite of 3 assessments, each rated from 0–3: stool frequency, rectal bleeding, and PGA
Modified Mayo Clinic Score (mMCS)	mMCS (0–9) is a composite of 3 assessments, each rated from 0–3: stool frequency, rectal bleeding, and endoscopy
MCS Remission	MCS ≤ 2 with individual subscores ≤1 and a rectal bleeding subscore of 0
MCS Clinical Remission	MCS ≤ 2 with individual subscores ≤ 1
MCS Clinical Response	MCS with ≥ 3-point decrease and 30% reduction from baseline as well as ≥ 1-point decrease in rectal bleeding subscore or an absolute rectal bleeding score of 0 or 1
pMCS Remission	pMCS ≤ 2, with, rectal bleeding score of 0, PGA 0–1 and stool frequency subscore 0–1
mMCS Remission (excluding friability from ES = 1)	mMCS ≤ 2, with, rectal bleeding score of 0, endoscopy 0–1 and stool frequency subscore 0–1 (Friability must be Absent for ES = 1)
mMCS Remission (including friability in ES = 1)	mMCS ≤ 2, with, rectal bleeding score of 0, endoscopy 0–1 and stool frequency subscore 0–1
pMCS Clinical Remission	pMCS ≤ 2, with, rectal bleeding score of 0–1, PGA 0–1 and stool frequency subscore 0–1
mMCS Clinical Remission (excluding friability from ES = 1)	mMCS ≤ 2, with, rectal bleeding score of 0–1, endoscopy 0–1 and stool frequency subscore 0–1 (Friability must be Absent for ES = 1)
mMCS Clinical Remission (including friability in ES = 1)	mMCS ≤ 2, with, rectal bleeding score of 0–1, endoscopy 0–1 and stool frequency subscore 0–1
pMCS Clinical Response	A decrease in the pMCS of at least 2 points and at least 30% improvement from baseline, with an accompanying decrease in the rectal bleeding score by at least one point or an absolute rectal bleeding score of 0 or 1.
mMCS Clinical Response (excluding friability from ES = 1)	A decrease in the mMCS of at least 2 points and at least 30% decrease (improvement) from baseline, with an accompanying decrease in the rectal bleeding score by at least one point or an absolute rectal bleeding score of 0 or 1. (Friability must be Absent for ES = 1)
mMCS Clinical Response (including friability in ES = 1)	A decrease in the mMCS of at least 2 points and at least 30% decrease (improvement) from baseline, with an accompanying decrease in the rectal bleeding score by at least one point or an absolute rectal bleeding score of 0 or 1.
Improvement in endoscopic appearance of the mucosa (including friability in ES = 1)	Endoscopy subscore ≤ 1

Endoscopic Remission (including friability in ES = 1)	Endoscopy subscore = 0
Improvement in endoscopic appearance of the mucosa (excluding friability from ES = 1)	Endoscopy subscore ≤ 1 (Friability must be Absent for ES = 1)

ES= Endoscopic Subscore; PGA = Physician's Global Assessment.
Endoscopy Score=1 includes mild friability criteria unless identified in the description of an endpoint above.

4.4.2.2 Histologic Endpoints

For each patient, scanned images of hematoxylin and eosin stained slides of FFPE sigmoid colon biopsies are assessed by the same pathologist from among a small pool of central readers. Details of the reading process are included in Section 3.2. Each slide image is evaluated using two histologic scoring systems: NHI (Appendix 4) and Geboes Grading Scale (Appendix 5). A third score, the RHI, is derived from selected components of Geboes Grading Scale (Appendix 5).

All histologic endpoints will be evaluated only on patients who have documented neutrophilic inflammation at baseline. Neutrophilic Inflammation will be defined using the scoring system used within the analysis. The main analysis of each histologic endpoint will be conducted on the basis of NHI scoring system, and sensitivity analyses will be conducted using RHI and Geboes Grading Scale (Table 4). This data was not collected in the GA29103 study.

Table 4 Histologic Endpoints

Feature or Outcome	Description	Definition for each scoring system		
		NHI	RHI	Geboes Grading Scale
	Indication of neutrophilic inflammation	NHI > 1	RHI > 3	Geboes 2B.1–2B.3 or 3.1–3.3 or 4.1–4.3 or 5.1–5.3
Histologic remission	Resolution of neutrophilic inflammation	NHI ≤1	RHI ≤3 and Geboes 2B.0 and 3.0	Geboes 2B.0 and 3.0, and 4.0, and 5.0

NHI=Nancy Histological Index; RHI=Robarts Histopathological Index.

4.4.2.3 UC-PRO

The UC-PRO questionnaire is collected in the e-diary and completed by patients for at least 9–12 consecutive days prior to a study visit as per the Schedule of Assessments in the protocol. The UC-PRO is being reported in three domains; two domains are key endpoints and reported as UC-PRO Signs and Symptoms. This data was not collected in the GA29103 study.

4.4.2.3.1 Functional and Bowel Domain

Two domain scores computed for the UC-PRO Signs and Symptoms:

- Functional Symptoms
- Bowel Movement Signs and Symptoms

There is no single total score. The questions contributing to each domain are shown in Appendix 6. The functional domain score ranges from 0–12, the bowel domain score ranges from 0–27, with a higher score indicating a worse disease state. The responder definition cutoffs for the Functional and Bowel domains will be pre-specified in the Data Analysis Plan (DAP).

The daily scores contributing to the UC-PRO/SS calculation for a visit will be selected as: most recent 7 daily scores available prior to but not including a visit. (Note: a minimum of 4 days is required.)

For each item in the questionnaire, a score will be calculated for a visit by taking the average of the most recent 7 daily scores available. The domain score for a visit will then be determined, taken as the sum of the (averaged) items for each question.

The endpoint using the UC-PRO/SS is the change from baseline at Week X in UC-PRO/SS Domain as assessed by UC-PRO/SS measure.

$$\textbf{Change from Baseline at Week X} = \textbf{Week X UCPRO/SS Domain Score} - \textbf{Baseline UCPRO/SS Domain Score}$$

A MMRM (Mixed Model Repeated Measures) analysis will be performed to assess the change from baseline in UC-PRO domains at Week X and will include the fixed categorical effects of treatment, visit, study stratification factors, and treatment-by-visit interaction, and the continuous covariates of the baseline continuous UC-PRO domain and baseline UC-PRO domain-by-visit interaction. An unstructured covariance matrix will be used to model the within patient errors within the MMRM.

4.4.2.3.2 Systemic Symptoms Domain

The systemic symptoms domain collected by the UC-PRO tool ranges from 0-20 with a higher score indicating a worse disease state. Details of the score are in Appendix 6. The scores are summarized using the same methodology used for functional and bowel domain in Section 4.4.2.3.1.

4.4.2.4 Inflammatory Bowel Disease Questionnaire

The IBDQ is a 32-item questionnaire containing four domains: bowel symptoms (10 items), systemic symptoms (5 items), emotional function (12 items), and social function (5 items).

An overall total IBDQ score will be computed by summing the individual 32-item scores. The range for the IBDQ total score is 32–224, with higher scores denoting better health-related quality of life.

The IBDQ questionnaire is administered once per visit as per the study Schedule of Assessments, and is completed by the patient at the clinic.

The change from baseline at Week X in total IBDQ

$$\textbf{Change from Baseline at Week X} = \textbf{Week X IBDQ} - \textbf{Baseline IBDQ}$$

This is analyzed by an ANCOVA model with the factors used at randomization into the Induction/Maintenance Phases as stratification variables, and the baseline IBDQ score used as a covariate.

4.4.3 Subgroup Analyses

The following subgroup analyses will be conducted on the primary endpoints for all studies. The subgroup categories are listed in Section 4.3. Study specific subgroups will be included in the study SAPs as required.

The primary endpoints will be summarized by the following subgroups using data collected in the clinical database:

- Baseline MCS
- Disease Location
- Age
- Gender
- Race
- IS use at Baseline
- CS use at Baseline
- Anti-drug antibodies (ADA) –ve/+ve (+ve transient or +ve persistent status)

Additional subgroups analysis will be conducted as appropriate.

4.5 PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Pharmacokinetic Analyses

Serum concentration at various times during Induction and Maintenance Phases will be listed and summarized by descriptive summary statistics including means, geometric means, ranges, SD, and coefficients of variation.

Individual and mean concentration versus time data will be tabulated and plotted if more than 2 time points are available.

The pharmacokinetic (PK) data from each individual study may be combined with data from other etrolizumab studies to perform a population PK analysis. Population typical value of PK parameters will be estimated for the entire study population, along with estimates of intra- and inter-patient variance and an estimate of random error. Individual patient parameter estimates will be computed using the post hoc analysis procedure. Impacts of covariates on relevant PK parameters will also be evaluated. A separate prospective analysis plan will be prepared, and the population PK analysis will be presented in a separate report for all studies, this will be separate from each study CSR.

Pharmacodynamic Analyses

The pharmacodynamic (PD) biomarker, sMAdCAM-1 absolute concentration and percentage change from baseline values will be listed and summarized by descriptive summary statistics at each time point including but not limited to means, SD, medians, and ranges. Analyses will be split by treatment group and cohort as appropriate.

Additional exploratory PK/PD analyses or modeling may be conducted as appropriate

4.6 BIOMARKER ANALYSIS

Additional biomarker strategies and analyses will be detailed in the Biomarker Analysis Plan.

4.7 SAFETY ANALYSES

The safety populations include all patients who received at least one dose of study drug, and patients will be grouped according to the treatment of the treatment arm they most frequently received. In addition, data will be listed for patients who do not receive the treatment they are assigned to in the safety population at any time point. All safety parameters will be summarized and presented in tables using the safety populations defined for each study. Patients who are not randomized but who receive study drug will be included in the safety population and summarized according to the therapy actually received. The safety data will be listed and summarized at the time of the primary analyses with use of all safety data available at the primary database snapshot. Additional summaries will be run once all patients have completed safety follow-up.

4.7.1 Adverse Events

Adverse events (AEs) will include all terms recorded on the AE Case Report Form (CRF) pages (except pregnancies). For each recorded AE, the term entered by the investigator describing the event (the “reported term”) will be assigned a standardized term (the “preferred term”) and assigned to a superclass term on the basis of the Medical Dictionary for Regulatory Activities (MedDRA) World Health Organization (WHO) dictionary of terms. All analyses of AE data will be performed using the preferred terms unless otherwise specified. For all summary tables, the AEs will be sorted by System Organ Class (in decreasing order of overall incidence) and then by preferred term (in decreasing order of overall incidence). In addition, separate summaries or listings will

be generated for SAEs, deaths, AEs leading to discontinuation of study drug, and adverse events of special interest (AESIs). In addition to summaries or listings, narratives will be provided for all deaths, SAE's, AE's leading to treatment discontinuation and pregnancies as well as for all serious infections, opportunistic infections and malignancies in the individual CSRs.

For the etrolizumab Phase III program, the AESIs (identified by the investigator using the eCRF tick box) are the following:

- Systemic hypersensitivity reactions and anaphylactic and anaphylactoid reaction. These will be further described using the MedDRA anaphylactic reaction SMQ (Standard MedDRA Query) algorithmic and Hypersensitivity SMQ narrow.
- Neurological signs, symptoms, and AEs that may suggest possible progressive multifocal leukoencephalopathy (PML) (see Appendices 5 and 6 of Protocol)
- Suspected transmission of an infectious agent by the study drug
- Cases of potential drug-induced liver injury that include an elevated alanine aminotransferase (ALT) or aspartate aminotransferase (AST) in combination with either an elevated bilirubin or clinical jaundice, as defined by Hy's Law

Specific AEs listed below will also be reported:

- Serious infections (including GI) – Events occurred in the MedDRA Infections and Infestations System Organ Class (SOC) using the primary coding events
- Opportunistic infections – Events occurred in the Sponsor-defined Adverse Event Group Terms
- Malignancies – Events that occur in the MedDRA Malignant and Unspecified Tumours SMQ (narrow).
- Injection site reactions – Events identified using the eCRF tick box indicating an Injection site reaction AND/OR events occurred in the MedDRA Injection Site Reaction High Level Term (HLT) using both primary and secondary coding.

Outputs will be summarized using the safety population split by treatment arm and cohort as appropriate.

Summaries of AEs will be generated to summarize the incidence of treatment-emergent AEs only. Treatment-emergent events are defined as any new AE reported or any worsening of an existing condition on or after the first dose of study drug. If the onset date of the AE is prior to the day of first dose, the AE will be considered treatment-emergent only if the most extreme intensity is greater than the initial intensity (i.e., the intensity for a given AE increases and its end date is on or after the date of the first dose). Non-treatment emergent AE's collected in the database will also be reported separately, for each individual CSRs.

For each treatment group, the incidence count for each AE preferred term will be defined as the number of patients reporting at least one treatment-emergent occurrence of the

event (multiple occurrences of the same AE in one patient will be counted only once). The proportion of patients with an AE will be calculated as the incidence count divided by the total number of patients in the population. Each table will also present the total number of AEs reported where multiple occurrences of the same AE in an individual are counted separately.

The rate per 100 patient years and 95% CIs will be summarized by treatment group for studies (GA28950, GA29102, GA29103, GA28948, and GA28949). Rates will be calculated for AEs, SAEs, and other AE grouping as appropriate. The rate of AEs per 100 patient years is calculated as:

$$= \frac{\textbf{Total Number of AEs}}{\textbf{Total Number of Patient Years at Risk}} \textbf{ X 100}$$

All summaries and listings of AEs will be based on the induction safety population or the maintenance safety population and listed in the List of Planned Outputs (LoPO).

4.7.2 Laboratory Data

Descriptive summaries of laboratory values at baseline and throughout the studies will be summarized by treatment arm.

Change from baseline will be analysed for the following parameters: Hematology (hemoglobin, hematocrit, platelet count, WBC count, lymphocytes, mean corpuscular volume), serum chemistries (sodium, potassium, chloride, bicarbonate, glucose, blood urea nitrogen, creatinine, calcium, total and direct bilirubin, total protein, albumin, alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase), CRP, and fecal calprotectin

Proportion of patients experiencing clinically significant changes relative to baseline will be summarized by treatment arm as appropriate for the parameters listed above.

Laboratory abnormalities and the patient's worst National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) grade during study will be summarized by treatment arm as appropriate for parameters listed above.

Elevated liver enzyme tests will be summarized by the following upper limit of normal (ULN) categories as these are indicators of severe liver injury

- ALT or AST $>3 \times$ ULN
- ALT or AST $>3 \times$ ULN and total bilirubin $>2 \times$ ULN as defined by Hy's law:

The number and percentage of patients with positive serum antibodies to etrolizumab (ADA) at baseline (prevalence) and post etrolizumab treatment at any time during the

study (incidence) will be tabulated by treatment arm and listed alongside the primary efficacy outcome.

All summaries and listings of laboratory data will be based on the safety populations and specified in the LoPO.

4.7.3 Vital Signs

Vital signs will be summarized using summary statistics and change from baseline proportion of patients experiencing clinically significant changes relative to baseline will be reported if appropriate.

All summaries and listings of vital signs data will be based on safety population and specified in LoPO.

4.7.4 Medical History

Medical history data collected in the e-CRF will be summarized using summary statistics, reporting the proportion of patients with at least one medical condition and the total number of medical conditions. The medical conditions will then be split out by type.

All summaries and listings of medical history will be based on safety population and specified in the LoPO.

4.7.5 Concomitant Medications

Concomitant medications include any medication being used at any time from first dose of study drug through to day of study discontinuation/completion, or medication being used at any time up to the start of study treatment. The data will be summarized, and report the total number of patients taking at least one medication, and total number of medications. Summaries will also be split by medication class and preferred medication.

All summaries and listings of concomitant medications will be based on safety population and specified in the LoPO.

4.8 MISSING DATA

The handling of intercurrent events explained in the individual studies will be applied before using the missing data handling approaches in this section. The handling of intercurrent events will occur in the order the intercurrent event occurs within the study.

To assess the robustness of the primary endpoint, a tipping point analysis will be conducted. The tipping point is defined as the difference in the number of missing events (i.e., remission) between the treatment groups that result in a change in the primary outcome conclusions (Yan et al. 2009). A two-dimensional plot will be produced for each primary comparison of etrolizumab (105 mg) vs. the comparator to evaluate where the tipping point lies.

The tipping point analysis will be used to assess the robustness of the primary endpoints within each study. The following groups of patients will always be considered nonremitters/non-responders within the tipping point analysis.

- Patients whose remitter/responder status can be calculated as a non-remitter/non-responder from their available sub scores, even when not all four MCS subscores are available.
- Patients who have received rescue therapy during the study.
- Patients who have discontinued study treatment early.

For all continuous endpoints, a MMRM model will be fitted to analyse UC-PRO domains assuming the data is missing at random. For missing SF/RB and IBDQ data single imputation worse observation carried forward (WOCF) post baseline will be applied to missing data, including when data is missing due to intercurrent event. An additional sensitivity analysis using observed case will be conducted, to assess the robustness of the analysis. All summary data will be reported using no imputation for missing data.

For any missing baseline data, no imputation of results data will be applied, and therefore, any endpoints requiring comparison back to baseline results will also be set to missing. To prevent a missing SF/RB subscore at baseline or post baseline which would lead to a missing MCS/mMCS/pMCS score, the anchor date used for selecting diary data to calculate RB/SF subscores can be imputed. If bowel prep date is missing, then endoscopy date can be used; if the endoscopy date is missing then the visit date can be used as the anchor date. If the PGA visit date is missing, the visit date will be imputed as Day 1+x weeks to allow RB and SF scores to be calculated.

No imputation will be applied for missing laboratory or vital signs data.

An AE with a completely missing, non-imputed start date will be assumed to be treatment-emergent, unless the AE has a complete, non-imputed end date that is prior to the date of the first dose.

All deaths will be reported, regardless of completeness of death date.

4.9 INTERIM ANALYSES

No interim analyses are planned for these studies.

5. REFERENCES

Geboes K, Riddell R, Öst A, et al. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. Gut. 2000;47(3):404–9.

Mosli MH, Feagan BG, Zou G, et al. Development and validation of a histological index for UC. Gut. 2017;66(1):50–8.

Marchal-Bressenot A, Scherl A, Salleron J, et al. A practical guide to assess the Nancy Histological Index for UC. Gut. 2016;65(11):1919–20.

Marchal-Bressenot A, Salleron J, Boulagnon-Rombi C, et al. Development and validation of the Nancy Histological Index for UC. Gut. 2017;66(1):43–9.

Yan X, Lee S, Li N. Missing data handling methods in medical device clinical trials. J Biopharm Stat. 2009;19(6):1085–98.

Appendix 1 Protocol Synopsis

See individual study SAPs for Protocol Synopsis.

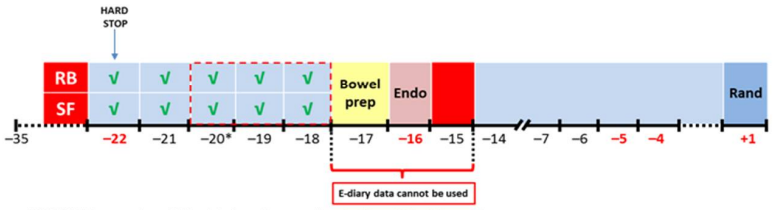
Appendix 2 Schedule of Assessments

See individual study SAPs for Schedule of Assessments.

Appendix 3 Mayo Clinic Score Measurement

Scenario 1: Sufficient e-Diary Data Available prior to Endoscopy at Day –16

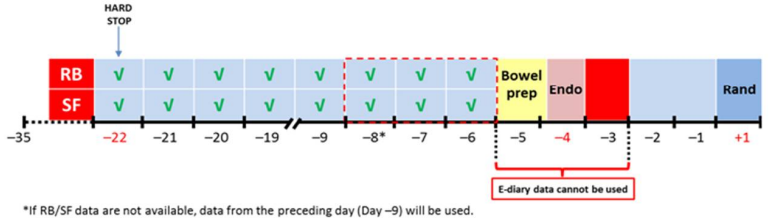
The e-diary data from the 3 consecutive days immediately preceding the bowel preparation day (Day –17 in this scenario) are used to derive RB and SF data for MCS calculation (Day –20 to Day –18, highlighted with dashed lines in the figure). If e-diary data from these days are missing, RB/SF data from the preceding days (Day –21 or Day –22) will be used. No RB/SF data can be obtained prior to Day –22.



*If RB/SF data are not available, data from the preceding day (Day –21) will be used.

Scenario 2: Sufficient e-Diary Data Available prior to Endoscopy at Day –4

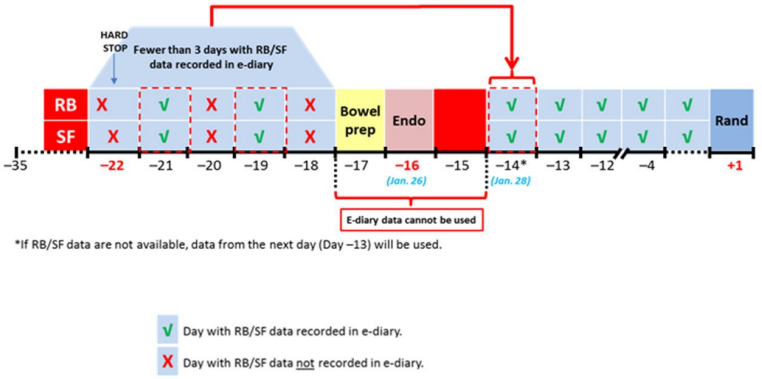
The e-diary data from the 3 consecutive days immediately preceding the bowel preparation day (Day –5 in this scenario) are used to derive RB and SF data for MCS calculation (Day –6 to Day –8, highlighted with dashed lines in the figure). If e-diary data from these days are missing, RB/SF data from the preceding days (Days –9 to Day –22) will be used. No RB/SF data can be obtained prior to Day –22.



*If RB/SF data are not available, data from the preceding day (Day –9) will be used.

Scenario 3: Insufficient e-Diary Data Available prior to Endoscopy at Day –16

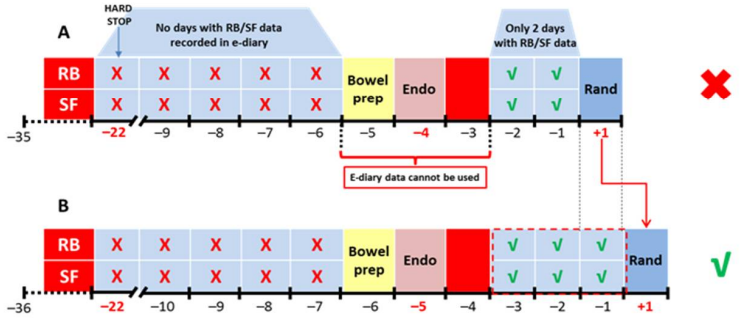
Only in cases where < 3 days of e-diary data are available prior to the bowel preparation day (Day –18 to Day –22 in this scenario), supplement with e-diary data starting 2 days after the endoscopy (e.g., January 28 if the endoscopy was performed on January 26). In the figure, the days highlighted with dashed lines can be used for MCS calculation.



*If RB/SF data are not available, data from the next day (Day –13) will be used.

Scenario 4: Insufficient e-Diary Data Available prior to Endoscopy at Day –4

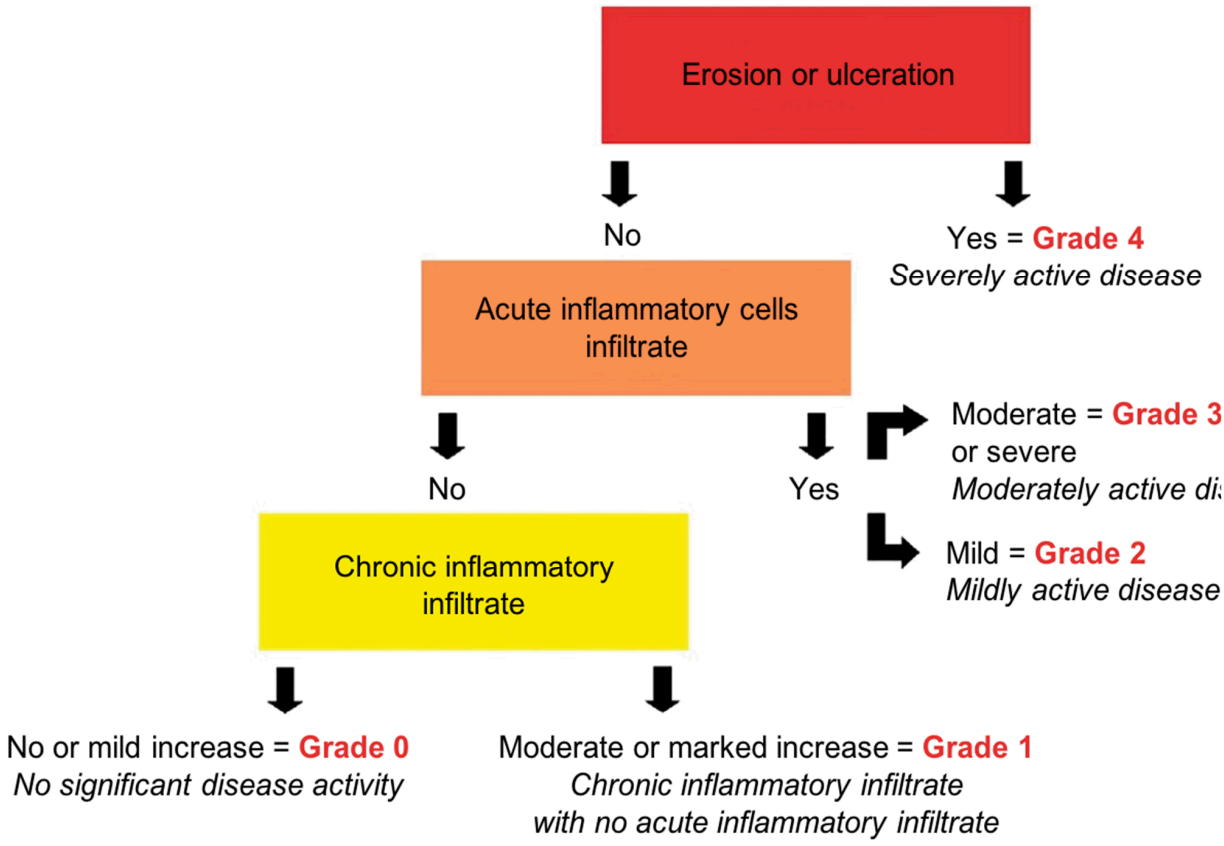
Only in cases where there are insufficient e-diary data (< 3 days total) available prior to the bowel preparation day (Day –6 to Day –22 in this scenario) and between the endoscopy and randomization (Day –2 and Day –1), the randomization visit must be delayed by at least 1 day by extending the screening period so sufficient data can be recorded in the e-diary (Day –1 in the extended screening period in Figure B). In the schema, the days highlighted with dashed lines can be used for MCS calculation.



Endo=day of endoscopy; MCS = Mayo Clinic Score; Rand=day of randomization; RB = rectal bleeding; SF = stool frequency.

Appendix 4 Nancy Histological Index

Nancy Histological Index (NHI): Scoring algorithm



Source: Adapted from Marchal-Bressenot et al. 2016, Marchal-Bressenot et al 2017.

Appendix 5 Geboes Grading Scale and Robarts Histopathological Index

Geboes Grading Scale: Grades and Subgrades

Grade 0: Structural (architectural change)

0.0 No abnormality

0.1 Mild abnormality

0.2 Mild-moderate diffuse/multifocal abnormalities

0.3 Severe diffuse/multifocal abnormalities

Grade 1: Chronic inflammatory infiltrate

1.0 No increase

1.1 Mild but unequivocal increase

1.2 Moderate increase

1.3 Marked increase

Grade 2A: Eosinophils in lamina propria

2A.0 No increase

2A.1 Mild but unequivocal increase

2A.2 Moderate increase

2A.3 Marked increase

Grade 2B: Neutrophils in lamina propria

2B.0 None

2B.1 Mild but unequivocal increase

2B.2 Moderate increase

2B.3 Marked increase

Grade 3: Neutrophils in epithelium

3.0 None

3.1 < 5% crypts involved

3.2 < 50% crypts involved

3.3 > 50% crypts involved

Grade 4: Crypt destruction

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

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

Source: Geboes et al. 2000.

Robarts Histopathological Index (RHI): Grade-Weighted Sum of Subgrades from Geboes

RHI score (0-33) = 1 × chronic inflammatory infiltrate (0-3)
+ 2 × neutrophils in lamina propria (0-3)
+ 3 × neutrophils epithelium (0-3)
+ 5 × erosion or ulceration (0-3)

Note: Erosion or ulceration component subgrade range of 0 to 3 is obtained by scoring Geboes subgrades as follows: 5.0 = 0, 5.1 or 5.2 = 1, 5.3 = 2, and 5.4 = 3.

Source: Mosli et al. 2017.

Appendix 6 Ulcerative Colitis Patient Reported Outcomes Signs and Symptoms

UC-PRO/SS Domain	Question	Response to Question
Functional Symptoms	In the past 24 hours, did you pass gas?	0=No 1=Rarely 2=Sometimes 3=Often 4=Very often
	In the past 24 hours, did you feel pain in your belly?	0=No 1=Mild 2=Moderate 3=Severe 4=Very Severe
	In the past 24 hours, did you feel bloating in your belly?	0=No 1=Mild 2=Moderate 3=Severe 4=Very Severe

Appendix 6 Ulcerative Colitis Patient Reported Outcomes Signs and Symptoms (contd.)

UC-PRO/SS Domain	Question	Response to Question
Bowel Movement Signs and Symptoms	In the past 24 hours, how many bowel movements did you have?	0=0, 1=1-2, 2=3-4, 3=5-6, 4=7-9, 5=10-12, 6=13-17, 7=18 or more
	In the past 24 hours, how often were your bowel movements mostly or completely liquid?	0=Never 1=Rarely 2=Sometimes 3=Often 4=Always
	In the past 24 hours, did you have blood in your bowel movements?	0=No 1=Rarely 2=Sometimes 3=Often 4=Always
	In the past 24 hours, did you have mucus (white material) in your bowel movements?	0=No 1=Rarely 2=Sometimes 3=Often 4=Always
	In the past 24 hours, did stool, blood, or liquid leak out before you reached a toilet?	0=No 1=Rarely 2=Sometimes 3=Often 4=Always
	In the past 24 hours, did you feel the need to have a bowel movement right away?	0=No 1=Mild 2=Moderate 3=Severe 4=Very Severe

Appendix 6 Ulcerative Colitis Patient Reported Outcomes Signs and Symptoms (contd.)

UC-PRO/SS Domain	Question	Response to Question
Systemic Symptoms	In the past 24 hours, did you feel pain in your knees, hips, and/or elbows?	0=No 1=Mild 2=Moderate 3=Severe 4=Very Severe
	In the past 24 hours, did you feel tired?	0=No 1=Mild 2=Moderate 3=Severe 4=Very Severe
	In the past 24 hours, did you lack an appetite?	0=No 1=Mild 2=Moderate 3=Severe 4=Very Severe
	In the past 24 hours, did you feel weak?	0=No 1=Mild 2=Moderate 3=Severe 4=Very Severe
	In the past 24 hours, did you feel thirsty?	0=No 1=Mild 2=Moderate 3=Severe 4=Very Severe